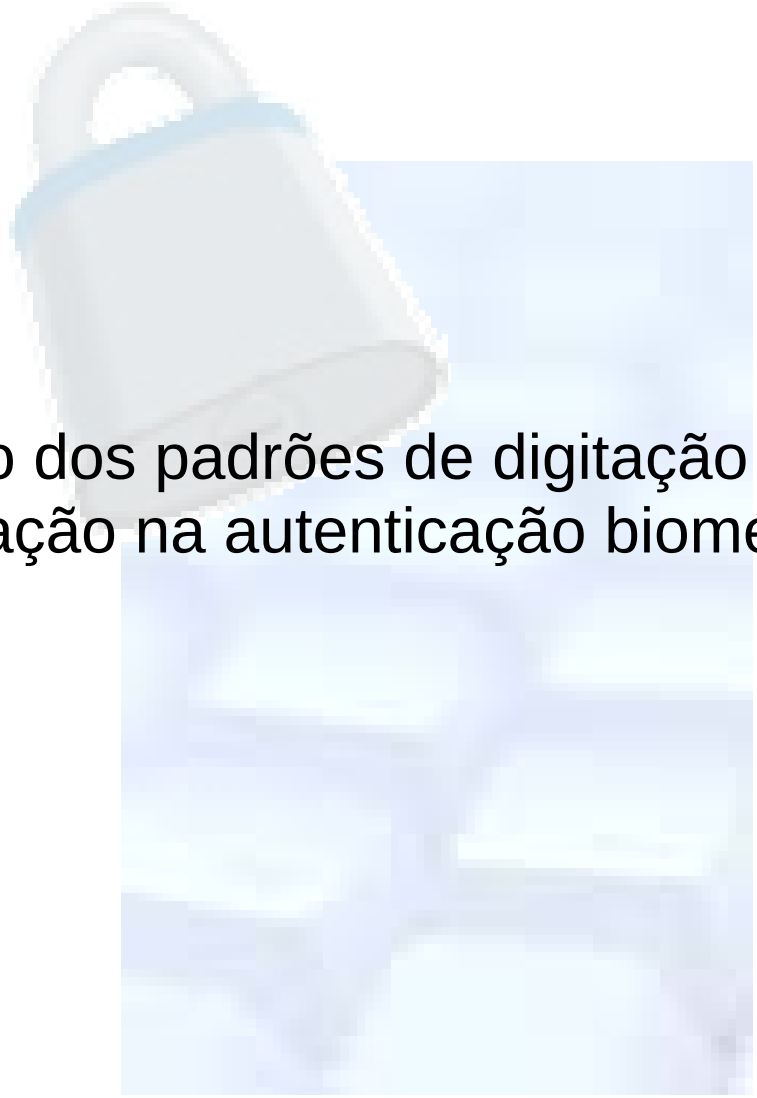
Universidade do Minho
Departamento de Sistemas de Informação



estudo dos padrões de digitação e sua aplicação na autenticação biométrica

Supervisor: Doutor Henrique Manuel Dinis dos Santos

"…We aren't going to be able to rely on passwords. Moving to biometric and smart cards is a wave that is coming and we see our leading customers doing this."

Bill Gates (em discurso no Microsoft IT Forum – Copenhaga, Novembro de 2004)

À Marizé e à Ana

Agradecimentos

O trabalho que originou esta dissertação não teria sido possível sem a colaboração daqueles que, voluntariamente, se dispuseram a apoiar-me por diversas formas. Esta é a minha oportunidade para lhes agradecer. Assim, ficam aqui expressos os meus agradecimentos:

- Aos alunos do curso EFA da Sol-do-Ave, ao Dr. Nuno Casalta, à Eng.ª Jorgete Pires e aos alunos de Sistemas de Computação (2003/2004) pela sua disponibilidade para ceder os dados biométricos,

- Aos alunos de Sistemas de Computação I (2004/2005) pela sua disponibilidade para responder aos questionários sobre práticas de segurança em Sistemas de Informação,

- A todos os voluntários que, anonimamente, participaram nas recolhas de dados efectuadas,

- À minha mãe pelo apoio financeiro,

- À Marizé e à Ana pela compreensão, apoio e inspiração.

Por último, mas não menos importante, quero agradecer ao Doutor Henrique Santos pela orientação, apoio e motivação.

Resumo

A segurança dos Sistemas de Informação (SI) é uma disciplina que atravessa horizontalmente diversas actividades das organizações, podendo afectar significativamente (sobretudo a falta de segurança) o seu desempenho. Neste contexto, a segurança levanta inúmeros desafios, como o da identificação e autenticação dos indivíduos perante o SI.

De uma forma genérica, um utilizador pode ser autenticado por algo que ele tem (um dispositivo), por algo que ele sabe (uma palavra passe), ou por alguma característica intrínseca. Várias técnicas surgiram para suportar estes tipos de autenticação, isoladamente e em conjunto e cada uma delas tem, naturalmente, virtudes e inconvenientes. Mas atendendo à sua potencial mais valia, as técnicas chamadas biométricas, que procuram utilizar características do indivíduo, têm vindo a evidenciar uma notável evolução. No entanto, a adopção destas tecnologias é travada pela desconfiança dos utilizadores quanto à utilização da sua informação privada e pelo receio de agressões à integridade física, por parte de algumas dessas tecnologias. O primeiro grupo de limitações é suavizado recorrendo a uma tecnologia de armazenamento e processamento da informação biométrica em *SmartCards*, sob controlo do próprio utilizador. Já o segundo é mais difícil de garantir atendendo à natureza intrusiva dos leitores biométricos. Mas existem técnicas biométricas pouco intrusivas. A chamada *Keystroke Dynamics*, ou dinâmica de digitação, é uma técnica que procura medir a forma como o utilizador usa o teclado. Alguns algoritmos foram propostos para este efeito, medindo tempos de latência entre teclas e utilizando diversos métodos para confrontar amostras de autenticação contra um padrão. No entanto, os resultados têm ficado aquém do que é necessário para uma utilização generalizada.

Nesta dissertação apresenta-se um novo algoritmo de *Keystroke Dynamics* desenvolvido. Procurou-se optimizar o modelo estatístico por forma a melhorar a sua precisão e gerando um algoritmo computacionalmente simples, garantido a possibilidade da sua execução nos limitados recursos de um *SmartCard*. Os testes realizados demonstram um melhor desempenho face aos restantes algoritmos e, apesar da ausência de *benchmarks*, é firme convicção do autor, dados os resultados obtidos, que o algoritmo proposto é de enorme utilidade para a implementação de medidas preventivas, relativas à identificação/autenticação, nas políticas de segurança para os Sistemas de Informação.

Abstract

The security of the Information Systems (IS) is a field of knowledge that crosses several activities inside an organization and that is able to affect significantly (especially the lack of it) its performance. In this context, security raises a number of issues, like identification and authentication of the individuals towards the IS.

In a generic way, a user can be authenticated by something he has (a device), by something he knows (a password), or by something he his. Several techniques were developed to match these kinds of authentication, isolated or together and each one with virtues and defects, naturally. But given their potentiality, the so-called biometric techniques, that try to use some intrinsic characteristic of the user, have presented a remarkable progress. Nevertheless, the adoption of these technologies is slowed by the suspicious of the users about the way their private information is going to be used and by the fear of damage to their physical integrity, by some of those technologies. The first group of limitations is softened by the use of a technology that allows the biometric data to be stored and processed inside a SmartCard, under the control of the user. The second group is harder to guarantee, considering the intrusive nature of most of the biometric readers. But there are some biometric techniques that are not intrusive. The so-called Keystroke Dynamics is a technique that tries to measure the way that a user types on the keyboard. Some algorithms were proposed for that matter, measuring the latency times between keystrokes and using different methods to compare the authentication samples with a pattern. But the results have been weaker then the ones needed for a wide use.

In this work a new algorithm of Keystroke Dynamics is presented. It was obtained by an optimization of the statistical model to achieve better precision levels and to generate an algorithm with few hardware requirements, ensuring the possibility of being executed in the limited resources of a SmartCard. The tests made show a better performance then the other similar algorithms and, despite the lack of benchmarks, the author is strongly convinced that the proposed algorithm is a major contribution for the implementation of preventive measures, regarding authentication and/or identification, in the information Systems security policies.

Índice

Introdução

A segurança dos Sistemas de Informação é uma disciplina que atravessa horizontalmente diversas actividades das organizações, podendo afectar significativamente (sobretudo a falta de segurança) o seu desempenho. Desde questões tecnológicas até questões culturais e comportamentais, é possível encontrar diversos trabalhos que procuram responder a velhos e novos desafios que se levantam, provenientes de novos modelos de organização e, sobretudo, de uma evolução tecnológica notável.

Na base da segurança dos SI está o problema de estabelecer uma associação entre um indivíduo e uma identidade, o qual pode ser dividido em duas grandes áreas: autenticação e identificação. *Autenticação* refere-se ao problema de confirmar ou negar uma alegada identidade de um indivíduo; enquanto *identificação* refere-se ao problema de estabelecer a identidade, desconhecida à partida, de um indivíduo [Thian, 2001]. O âmbito desta dissertação é a autenticação de um sujeito ligado, directa ou indirectamente, ao indivíduo ou organização que pretende confirmar a sua identidade.

A autenticação fraudulenta pode acarretar custos para uma organização. Muitos exemplos podem ser apresentados como o acesso não autorizado à contabilidade de uma empresa por alguém que obteve a palavra passe do contabilista, o acesso a um laboratório de alta segurança, ou simplesmente o acesso a informação estratégica por alguém que se faz passar por um utilizador legítimo.

A procura de um método de autenticação tem sido objecto de investigação intensa envolvendo, tradicionalmente, sistemas que têm a ver com a partilha de um segredo entre utilizador e objecto de segurança. Um dos problemas deste método é a transmissibilidade do segredo que, como qualquer outro, pode ser cedido (voluntariamente ou não) por quem o conheça a terceiros. Outro problema deste método é a necessidade de armazenamento ou memorização do segredo. Quando o segredo é armazenado, naturalmente herdamos o conjunto de vulnerabilidades que o(s) sistema(s) de armazenamento evidencia(m). Quando o segredo é memorizado pode ser esquecido, o que normalmente leva à escolha de segredos simples que facilitem a respectiva memorização, com consequências graves para as vulnerabilidades associadas.

A resposta a estas questões pode passar por soluções que permitam complementar os métodos existentes de autenticação com um local de armazenamento seguro e algum factor de identificação inerente ao sujeito autenticado, que dispense a criação arbitrária de segredos. É assim que surgem, no contexto da autenticação, os SmartCards e a autenticação biométrica.

Cada uma das técnicas de identificação e autenticação desenvolvidas têm, naturalmente, virtudes e inconvenientes, mas atendendo à sua potencial mais valia, as técnicas chamadas biométricas, que procuram utilizar características do indivíduo, têm vindo a evidenciar uma notável evolução. No entanto, a adopção destas tecnologias é travada pela desconfiança dos utilizadores quanto à utilização da sua informação privada e pelo receio de agressões à integridade física, por parte de algumas dessas tecnologias. O primeiro grupo de limitações é suavizado recorrendo a uma tecnologia de armazenamento e processamento da informação biométrica em *SmartCards*, sob controlo do próprio utilizador. Já o segundo é mais difícil de garantir atendendo à natureza intrusiva dos leitores das características biométricas.

Mas existem técnicas biométricas pouco intrusivas. A chamada *Keystroke Dynamics*, ou dinâmica de digitação, é uma técnica que procura medir a forma como o utilizador usa o teclado. Alguns algoritmos foram propostos para este efeito, medindo tempos de latência entre teclas e utilizando diversos métodos estatísticos para confrontar amostras de autenticação contra um padrão. No entanto, os resultados publicados têm ficado aquém do que é necessário para uma utilização generalizada.

Este trabalho pretendeu desenvolver um novo algoritmo de *Keystroke Dynamics*. optimizando o modelo estatístico por forma a melhorar a sua precisão, ao mesmo tempo que tendo a preocupação de gerar um algoritmo computacionalmente simples, de forma a garantir a possibilidade da sua execução nos limitados recursos de um *SmartCard*. Devido à ausência de *benchmarks* o algoritmo deve ser testado, usando o mesmo grupo de dados, contra outros algoritmos existentes para permitir uma avaliação comparativa.

Para enquadrar devidamente o trabalho que é descrito neste documento, os capítulos 1, 2 e 3 desta dissertação destinam-se, respectivamente, a apresentar as tecnologias relacionadas com SmartCards (de modo sintético), a influência e as questões relacionadas com a introdução das tecnologias biométricas nas organizações e as tecnologias biométricas de autenticação/identificação.

O capítulo seguinte, capítulo 4, refere-se ao trabalho experimental desenvolvido: um novo algoritmo de autenticação através da dinâmica de digitação (*keystroke dynamics*), as hipóteses colocadas e os resultados obtidos na sua avaliação.

Por último, apresentam-se as conclusões e a proposta de trabalhos futuros (capítulo 5).

1 – SmartCards

O primeiro cartão de plástico foi emitido pelo Diners Club em 1950 e, no final dessa mesma década, a American Express e a Carte Banche lançaram também a sua proposta para esse tipo de cartões. O primeiro cartão de crédito, que viria a transformar-se no VISA, foi emitido pelo Bank of America algum tempo mais tarde, sendo seguido pelo Mastercard. Estes primeiros cartões só possuíam as informações que estavam gravadas no plástico. A International Air Transportation Association (IATA) desenvolveu nos anos 70 o primeiro cartão com banda magnética, permitindo o armazenamento de 80 caracteres alfanuméricos. Este tipo de cartões pode, quando a gravação é feita por processos ópticos, armazenar alguns megabytes de informação mas, em contrapartida, têm ainda um custo elevado. A evolução seguinte na tecnologia dos cartões foi o aparecimento de cartões de memória com *chip*, baseados em circuitos microelectrónicos. O primeiro passo nesse sentido foi dado, em 1974, pela Innovatron. Por seu lado, a Bull melhorou a tecnologia, produzindo o primeiro cartão com microprocessador, em 1979. Este cartão tinha o processador num *chip* separado, uma solução que se mostrou pouco segura, situação que se manteve até à década de 80 [Berta, 2000].

Genericamente, dizemos que um SmartCard é um cartão com as dimensões de um cartão de crédito, munido de um *chip* com ou sem microprocessador. As dimensões e os protocolos de comunicação de um SmartCard estão normalizadas pela norma ISO 7816 [ISO/IEC 7816], que também inclui especificações para cartões híbridos (*chip* mais banda magnética). Fora destas dimensões encontram-se os cartões SIM (Subscriber Identification Module), utilizados pelo sistema de comunicações móveis GSM (Global System Mobile Communication) e seus descendentes. No entanto, alguns autores consideram estes cartões como um tipo diferente de SmartCard, assim como alguns outros dispositivos munidos de um *chip* em forma de botão, de anel, ou de outro qualquer acessório. No contexto deste trabalho apenas consideramos os SmartCards equipados com processador.

Um SmartCard possui três tipos de memória: *Random Access Memory* – RAM – volátil; EEPROM – *Electric Erasable Programmable Read Only Memory* – apenas disponível para leitura em funcionamento normal, mas alterável recorrendo a um modo de funcionamento dedicado; e *Read Only Memory* – ROM – gravada no processo de

fabrico do cartão e não alterável. A capacidade de armazenamento e processamento de um SmartCard é reduzida e limita as suas funcionalidades, apesar de a sua configuração ter evoluído ao longo dos últimos anos, num processo semelhante à evolução dos pequenos computadores no final dos anos oitenta. A limitação imposta pelo hardware do SmartCard é habitualmente contornada através da repartição do processamento necessário, entre o cartão e o terminal a que ele vai ser ligado – *host* – que pode ser de diversos tipos, desde que esteja equipado com um CAD (*Card Acceptance Device*).

Existem diversas tecnologias para programar SmartCards, orientadas, frequentemente, para situações tipo e que de seguida são sumariamente descritas, com base nas compilações exaustivas efectuadas por [Chen 2000] e por [Berta 2000].

3.1 – GSM

GSM é o nome genérico de um conjunto de normas publicadas pelo European Telecommunications Standard Institute destinadas à utilização em sistemas de comunicações envolvendo sistemas telefónicos. Esta norma tem um nível de aceitação cada vez maior, não só na Europa onde surgiu, como na Ásia e, mais recentemente, no continente Americano. Esta especificação recorre à utilização de um tipo específico de SmartCard, já referido, o SIM card.

3.2 – Java Card

A empresa SUN desenvolveu a Java Card Virtual Machine, uma Java Virtual Machine (JVM) limitada, capaz de interpretar um subconjunto da linguagem Java e o protocolo de comandos específicos suportados pelos SmartCards. A especificação Java Card tem sido implementada em diversos SmartCards tradicionais, como o DelaRue's Galactic, o Bull's Odyssey e o CyberFlex da Schlumberger, e noutros dispositivos, com formas diversas que vão desde um anel (figura 1), até brincos colocados nas orelhas do gado para armazenar o histórico de vacinação. Outro dispositivo interessante que implementa a especificação Java Card é o Bull's SIM Rock'n Tree, um cartão SIM capaz de interpretar e processar as mesmas aplicações que um SmartCard e com funções GSM.

Esta especificação, que define um ambiente onde os programas entram e saem livremente do cartão mas onde (por questões de segurança) não existe

interoperabilidade entre eles, sofre particularmente com as limitações de hardware, uma vez que elas impedem o aproveitamento do facto de o Java ser orientado a objectos.

Figura 1 - Java Ring e respectivo CAD

3.3 – EMV

A especificação EMV – Europay, Mastercard & VISA – é baseada na norma ISO 7816 [ISO/IEC 7816] e foi desenvolvida de modo a incluir extensões adequadas às necessidades das empresas financeiras. A primeira versão desta especificação foi apresentada em Junho de 1996 e a VISA está a preparar a migração, na Europa, do seu sistema para esta plataforma. A VISA espera ter a maioria dos terminais do seu sistema europeu preparados para aceitar cartões com *chip,* no início de 2005 [VISA, 2004].

3.4 – Open Platform

A especificação OP (*Open Platform*) foi inicialmente desenvolvida pela VISA e, mais tarde, transferida para a GlobalPlatform. Esta especificação tinha como objectivo a uniformização das implementações em tecnologias ligadas a SmartCards. A especificação OP exige que os leitores sejam compatíveis com as normas ISO e com a especificação EMV (também desenvolvida pela VISA) e define características que uma aplicação deverá possuir para ser uma tecnologia independente do fabricante do leitor e dos cartões.

3.5 – PC/SC

A especificação PC/SC (*Personal Computer/SmartCard*) propõe uma arquitectura para a utilização de *SmartCards* em computadores pessoais. Foi desenvolvida pelo PC/SC *Workgroup,* fundado em 1996 por empresas líderes do

mercado de computadores pessoais e de *SmartCards*: Bull CP8, Gemplus, Hewlett-Packard, IBM, Microsoft, Schlumberger, Siemens Nixdorf, Sun Microsystems, Toshiba e Verifone. De acordo com esta especificação, divulgada em 1997, os programas executados no sistema anfitrião (um computador pessoal) são construídos sobre um ou mais fornecedores de serviços e um gestor de recursos. O fornecedor de serviços transforma as especificidades de cada fabricante, tornando-as transparentes para o utilizador. O gestor de recursos, como o nome indica, gere os recursos necessários para o acesso do sistema ao CAD – aqui denominado *Interface Device* – e, a partir daí, a um cartão. Este gestor de recursos deverá: detectar os leitores existentes e, consequentemente, os tipos de cartões disponíveis; gerir os acessos concorrentes a um cartão e detectar a inserção e remoção dos cartões de modo a identificar e informar as aplicações dos cartões e serviços disponíveis a cada instante. A especificação PC/SC está essencialmente dirigida para o sistema Windows e, neste Sistema Operativo, qualquer aplicação desenvolvida segundo o *Opencard Framework* consegue aceder ao dispositivo leitor de cartões através do gestor de recursos do PC/SC.

3.6 – Open Card

O consórcio OpenCard (www.opencard.org) foi criado pelas principais empresas ligadas aos SmartCards, com o objectivo de criar plataformas normalizadas que permitam a interoperabilidade de aplicações independentemente do produtor do cartão/leitor. Apesar de não haver ainda uma norma que permita programar para qualquer aparelho, é já possível, em muitos casos, programar o acesso ao cartão sem a preocupação de programar explicitamente os acessos à porta física. Ao atingir níveis de abstracção mais elevados, com a consequente facilidade de integração e programação, é admissível uma clara afirmação desta tecnologia. De notar ainda, que esta especificação foi desenhada tendo em vista o funcionamento em redes informáticas e as API (*Application Program Interface*) fornecidas pelo consórcio estão implementadas em linguagem Java.

2 – A tecnologia biométrica nas organizações

No final do século passado, com a proliferação das tecnologias informáticas (nomeadamente o PC) e o avanço dos estudos sobre biometrias, tornou-se viável a implementação de autenticação por recurso a características biométricas dos indivíduos. No entanto, estas soluções, além das dificuldades técnicas, acarretavam algumas dificuldades de carácter social.

Países como a Austrália, Canadá, Estados Unidos e Nova Zelândia testemunharam uma inquietação pública quanto aos esquemas de identificação. Entre os vários receios citados incluem-se [Davies 1994]:

- Que as pessoas sejam desumanizadas ao serem reduzidas a códigos,
- Que o sistema potencie o poder sobre os indivíduos por parte de determinadas organizações e até do estado,
- Que a identificação de alta integração envolva a inversão da apropriada relação entre o cidadão e o estado,
- Que o sistema seja conduzido por uma burocracia tecnologicamente assistida, ao invés de por governos eleitos,
- Que isenções e excepções existam para organizações e indivíduos poderosos,
- Que estes esquemas de identificação sejam os mecanismos previstos em profecias religiosas como, por exemplo, a *Marca da Besta*.

Com a generalização de equipamentos leitores de características biométricas e com a sua divulgação em filmes de grande sucesso, o cidadão comum encara hoje a autenticação biométrica como algo que lhe é familiar, embora a associe em grande parte à ficção científica. Os resultados de um inquérito realizado em Novembro de 2004 pela Epaynews [Epaynews, 2004] indicavam que 37% dos inquiridos afirmaram preferir um sistema biométrico para a sua autenticação ao realizar pagamentos com cartões, enquanto que apenas 9% prefere a verificação da assinatura (figura 2). Só os códigos PIN se aproximam deste nível de confiança.

Por outro lado, o medo provocado pelo terrorismo, nomeadamente o atentado de 11 de Setembro de 2001 ao *World Trade Center*, levou os governos a aumentar os gastos em aquisição de tecnologias biométricas para autenticação de indivíduos na sua qualidade de cidadãos ou de funcionários [IBG, 2003]. A tecnologia biométrica tem

sido utilizada pelos governos ocidentais para reforçar os métodos de combate ao terrorismo. Os Estados Unidos da América (EUA) decidiram fotografar (com o

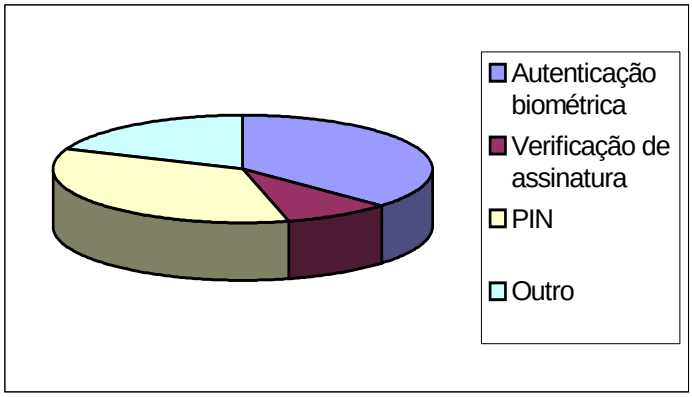


Figura 2 - Métodos de autenticação preferidos pelos utilizadores no uso de cartões de crédito. Fonte: epaynews.com.

objectivo de utilizar as imagens em sistemas de reconhecimento facial) e recolher as impressões digitais (electronicamente) dos visitantes estrangeiros que entrem no país com um visto no seu passaporte. Por outro lado, os EUA exigiram aos países com acordos que dispensam os seus cidadãos de vistos em estadias curtas, a criação de um programa tendo em vista a introdução de dados biométricos nos seus passaportes, até 26 de Outubro de 2004. Os visitantes oriundos dos países que não conseguiram cumprir este prazo sujeitam-se agora, à entrada no país, à introdução de dados biométricos (duas imagens digitais do dedo indicador e uma fotografia digital) no sistema norte-americano [U. S. Department of Homeland Security, 2004] [U. S. Department of State, 2004/886]. Também o Reino Unido passou a recolher dados biométricos (impressão digital) dos cidadãos da Etiópia, do Djibuti, da Eritréia, da Tanzânia e do Uganda, que solicitem um visto de permanência, bem como a todos os indivíduos africanos que viagem com o estatuto de refugiados. Além disso, o Reino Unido iniciou testes com o objectivo de introduzir dados biométricos nos bilhetes de identidade dos seus cidadãos, nomeadamente relativos à impressão digital e ao padrão da íris. No entanto, vários grupos de defesa dos direitos civis têm-se manifestado contra a introdução da biometria no controlo de fronteiras. Numa carta enviada à *International Civil Aviation Organization* (ICAO), a *Privacy International*, a *Statewatch*, a *European Digital Rights*, a *American Civil Liberties Union* e outras associações, alegam que a introdução da tecnologia biométrica tem um efeito na perda de privacidade e de direitos civis que é

desproporcional às vantagens de segurança que proporciona. Estes grupos criticam ainda a adopção, por este organismo, do reconhecimento facial como norma, invocando as altas taxas de erro desta tecnologia [Privacy International, Março 2004]. Os seus argumentos são apresentados por aquelas instituições, em carta aberta, ao Parlamento Europeu [Privacy International, Novembro 2004].

Entretanto, algumas empresas começam a utilizar a tecnologia biométrica como forma de apresentar produtos e serviços tradicionais de uma forma mais segura. Por exemplo, a marca japonesa NTT DoCoMo lançou em 2003 o primeiro telemóvel accionado biometricamente (produzido pela Fujitsu) e vendeu 700.000 unidades em menos de três meses. Em Portugal, a Galp Energia tem disponível em diversos postos de abastecimento de combustível o pagamento através da impressão digital, dispensando o cliente da apresentação do cartão multibanco.

O sector bancário começa também a utilizar a biometria para acesso aos seus produtos. Segundo o [Biometric Watch$_{TM}$, 2004], o Bancafe Bank, uma das principais instituições financeiras da Colômbia, está a incorporar sistemas de leitura de impressão digital nas suas máquinas ATM (*Automated Teller Machine*). A impressão digital é usada em conjunto com um código numérico secreto dispensando os seus clientes do uso do cartão bancário. Em Israel, o Leumi Bank of Israel adoptou um sistema de assinatura recolhida digitalmente para reduzir o risco de fraude na utilização de cheques.

A tecnologia biométrica é também utilizada para fins que não estão ligados à segurança. A Rhinowatch, um grupo com sede em Portugal dedicado à preservação da vida animal, utiliza imagens digitais de pegadas dos animais para identificar na Namíbia rinocerontes brancos e pretos. Este grupo adaptou agora o seu algoritmo para o aplicar aos tigres de Bengala, na Índia, em 2005. O grupo está a colaborar com outras associações para ajudar a proteger o tigre da Sibéria e o grande carnívoro mais ameaçado do mundo, o lince ibérico.

Em 2001 as biometrias eram utilizadas maioritariamente para controlo de acessos físicos (fig. 3), mas é de salientar que a aplicação destas tecnologias para controlo de acesso lógico é já significativa, representando 40% do total de utilizações destas tecnologias [Luis-García, 2003]. Neste âmbito, a biometria pode ajudar a ultrapassar as falhas de segurança inerentes à má utilização das palavras/frases passe. Num inquérito, realizado no âmbito desta dissertação, a sessenta pessoas que lidam com

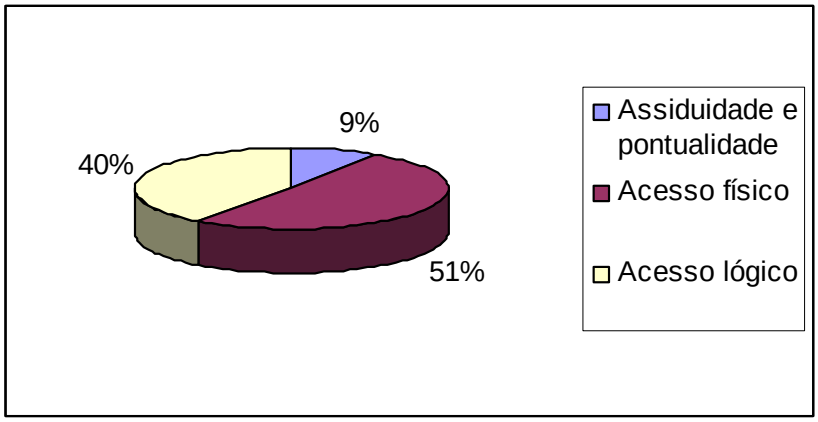


Figura 3 – Distribuição das tecnologias biométricas por aplicação (2001).
Fonte: [Luis-García, 2003]

sistemas protegidos por palavras/frases passe na sua vida pessoal e profissional, 74% dos inquiridos afirmaram que raramente alteram os seus códigos de acesso (figura 6), apesar de 52% dos inquiridos terem conhecimento de que eles são conhecidos por pelo menos uma pessoa (fig. 7). A figura 4 mostra a distribuição dos inquiridos por constituição das suas palavras/frases passe. Apenas 17% utiliza símbolos nos seus códigos, tornando-os, assim, mais vulneráveis.

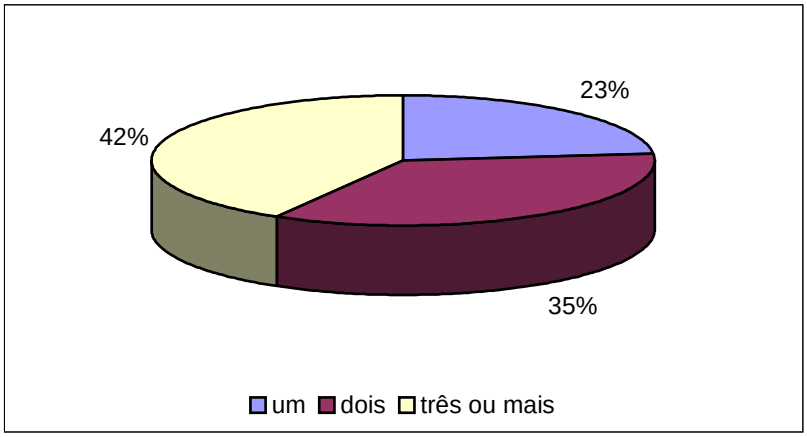


Figura 4 – Número de palavras passe usadas com frequência

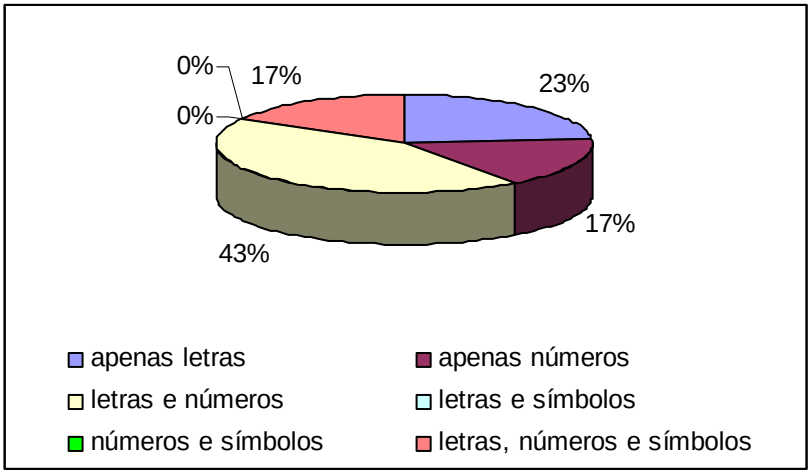


Figura 5 – Constituição das palavras/frase passe

A tecnologia biométrica pode também ajudar o utilizador a reduzir o número de palavras/frases passe que utiliza. Ainda com referência ao estudo anterior, apenas 23% dos inquiridos afirmaram utilizar, com frequência um só código secreto (figura 4). A utilização de palavras passe diferentes para acesso a diferentes serviços permite, em caso de divulgação de uma delas, manter seguro o acesso aos restantes. Com o uso de biometrias como reforço de segurança das palavras/frases passe, mesmo que esta seja tornada pública o acesso ao serviço continua protegido. Assim, um só código secreto é suficiente e, uma vez que é o único a ser memorizado, pode ser mais complexo.

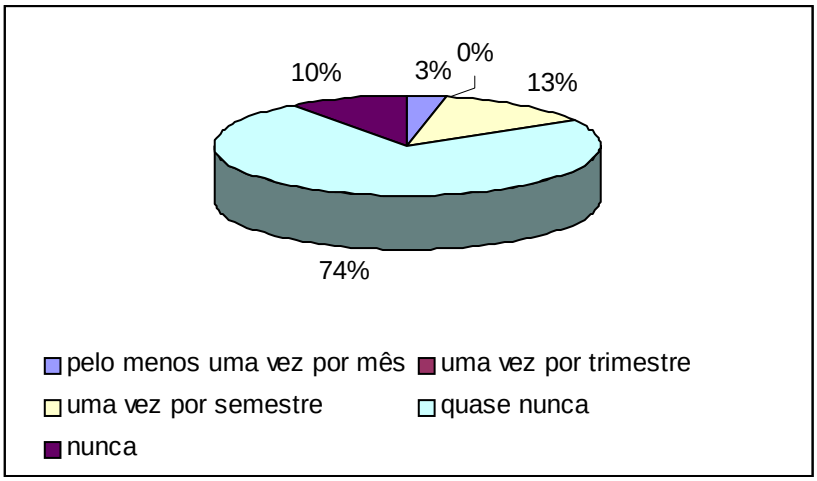


Figura 6 – Frequência de alteração de palavras/passe

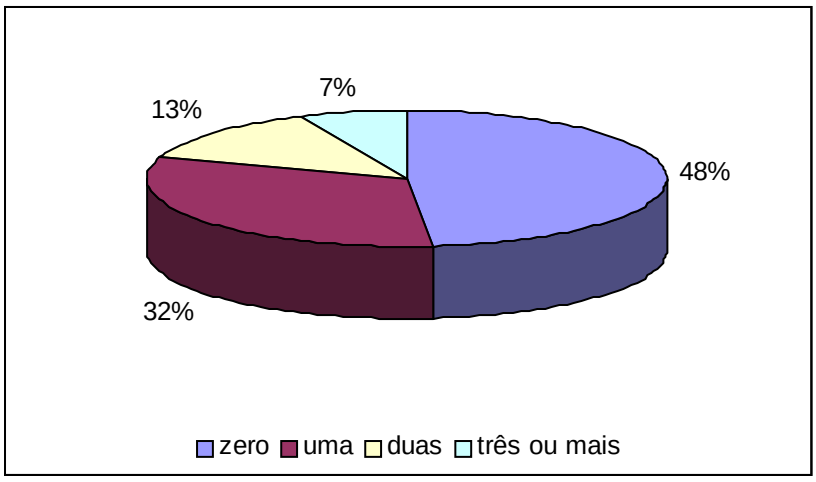


Figura 7 – Número de pessoas que o utilizador sabe que conhecem a(s) sua(s) palavra(s)-passe

Em 2002 o mercado das aplicações biométricas era dominado pelos sistemas de leitura de impressão digital (figura 8) mas a existência de sistemas biométricos com melhor precisão (a custos cada vez mais baixos) pode alterar esta situação num futuro próximo [Luis-García, 2003]. Outro factor que pode levar a uma inversão no domínio do mercado é a necessidade cada vez maior de identificar um indivíduo sem o seu conhecimento/consentimento, nomeadamente pelas forças policiais, em ambientes densamente ocupados como eventos desportivos, centros comerciais, estações de caminhos de ferro, aeroportos, etc. Assim, as tecnologias furtivas podem vir a assumir um protagonismo que hoje não detém.

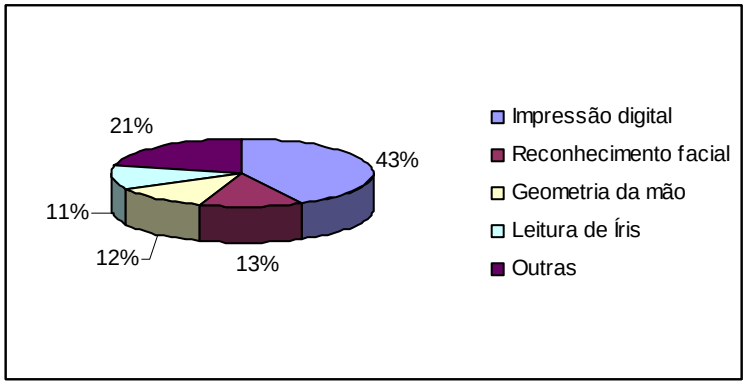


Figura 8– Distribuição do mercado por tecnologia biométrica (2002).
Fonte: [Luis-García, 2003]

Em Portugal é a Comissão Nacional de Protecção de Dados (CNPD) que tem a responsabilidade de autorizar a instalação de sistemas biométricos [Lei 67/98, 1998], com base nas evidências de garantia da protecção dos direitos daqueles que as utilizam. Devido à grande utilização destes sistemas para controlo de acessos e pontualidade, a CNPD tem disponível um documento [CNPD, 2004] com os princípios orientadores que devem reger a aplicação desta tecnologia, independentemente da obrigatoriedade de notificar a CNPD dos tratamentos de dados efectuados e de dispor de um parecer positivo desta instituição. Os próximos parágrafos baseiam-se nesse documento e pretendem sintetizar o contexto legal para a implementação de sistemas biométricos de controlo de assiduidade e pontualidade em Portugal, quer a nível social quer a nível técnico.

A utilização de sistemas biométricos no contexto de uma relação de trabalho deve ser precedida de um processo de esclarecimento e formação dos utentes de forma a obter a adesão voluntária dos trabalhadores e maximizar a eficácia do sistema. Só é exigível do trabalhador um dever de cooperação se existir um perfeito esclarecimento da forma como os dados recolhidos serão tratados e dos motivos que levam a entidade patronal a adoptar um sistema biométrico. Ainda assim, o trabalhador pode recusar o tratamento dos dados quando existirem "razões ponderosas e legítimas relacionadas com a sua situação particular". Além disso não é admissível o uso de tecnologias biométricas furtivas no contexto de uma relação de trabalho, uma vez que o titular do padrão biométrico tem o direito se saber se este se encontra armazenado e para que fins, bem como o direito de testar a validade desse padrão através da execução do processo de identificação e/ou identificação. Deve também existir um período de utilização experimental que permita avaliar o desempenho do sistema e deve ser possível ao trabalhador, de modo a satisfazer o disposto no artigo 17.º n.º4 do Código do Trabalho, a verificação do resultado do algoritmo biométrico sempre que o utilize, por exemplo através da apresentação num monitor da identidade identificada ou da existência de um sistema de luzes que confirme a correcta autenticação.

Uma vez que os padrões biométricos armazenados são uma representação digital da característica medida e, portanto, não permitem a duplicação ou reconstituição desta, está actualmente aceite pela CNPD que "a recolha de dados biométricos (...) não tem qualquer implicação com a integridade física do trabalhador, não afectando, igualmente, o seu direito à identidade pessoal e à intimidade da vida privada". Por forma a garantir

que assim seja, os processos de pedido de autorização de aplicação de sistemas biométricos apresentados à CNPD devem incluir uma declaração dos fabricantes de que não cedem às entidades que fornecem ou adquirem os equipamentos as chaves das representações digitais armazenadas.

Do ponto de vista técnico, os sistemas devem possuir um grau de fiabilidade suficiente para não comprometer a finalidade para que está a ser utilizado e não criar dificuldades acrescidas ao trabalhador, violando os seus direitos. Devido ao perigo para a privacidade da centralização de informações em bases de dados, não é admissível o relacionamento das tecnologias biométricas com outras como, por exemplo, a videovigilância, sem prejuízo da utilização de sistemas multimodais que recorram à avaliação de mais do que uma característica do trabalhador, de modo a aumentar a fiabilidade do processo. Apesar de as bases de dados constituídas serem um repositório de características inferidas a partir dos dados biométricos e não de dados biométricos em si, a CNPD recomenda que os padrões biométricos (em especial no caso da impressão digital) sejam armazenados em cartões transportados pelos utilizadores. Por último, deve ser referido que os dados biométricos de um utilizador devem obrigatoriamente ser eliminados no momento em que cesse a relação contratual ou em que o trabalhador mude de local de trabalho.

3 – As tecnologias biométricas

3.1 – Características gerais

O termo biometria deriva do grego *bios* (vida) + *metron* (medida) e, na autenticação, refere-se à utilização de características próprias de um indivíduo para proceder à sua autenticação e/ou identificação perante um SI de uma organização.

Existem hoje muitas características utilizadas, isoladamente ou em conjunto, para autenticar e/ou identificar um sujeito. As tecnologias biométricas são, normalmente, classificadas como comportamentais (por exemplo, reconhecimento de voz) ou físicas (por exemplo, leitura de retina), de acordo com a classificação das características avaliadas. Mas elas podem também ser classificadas como colaborativas, se exigem que o utilizador tenha conhecimento da sua existência e participe conscientemente no processo, ou como furtivas, se podem ser utilizadas sem o conhecimento daquele que é identificado ou autenticado [Magalhães, 2003]. Cada um dos métodos de autenticação pode ser avaliado através de vários parâmetros como o grau de fiabilidade, o nível de conforto, o nível de aceitação e o custo de implementação. [Liu et al. 2001].

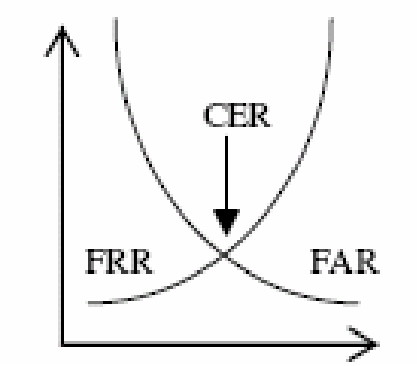


Figura 9 - Crossover Error Rate – CER

O grau de fiabilidade pode ser aferido tendo em atenção os valores FAR (*False Acceptance Rate* – Taxa de Falsas Aceitações) e o FRR (*False Rejection Rate* – Taxa de Falsas Rejeições). Infelizmente estas variáveis são mutuamente dependentes, não sendo possível minimizar ambas num mesmo algoritmo. Assim, normalmente procura-se o

ponto de equilíbrio (figura 9) a que chamamos CER (*Crossover Error Rate* – Taxa de Intersecção de Erros). Quanto mais baixo for o CER mais preciso é um sistema biométrico [Liu et al. 2001].

O nível de conforto é um padrão de certa forma subjectivo e está profundamente ligado ao público utilizador do sistema.

Outro padrão subjectivo é o nível de aceitação. De um modo geral o sistema é tanto melhor aceite pelos utilizadores quanto menos intrusivo for.

O custo de implementação é um factor fundamental e abrange diversos factores, alguns dos quais frequentemente descurados [Liu et al. 2001]:

- Hardware,
- Software,
- Integração com hardware/software existentes,
- Formação dos utilizadores,
- Pessoal de manutenção de Bases de Dados,
- Manutenção do sistema.

A escolha do(s) método(s) a utilizar depende da análise de risco que necessariamente deve ser feita, relativamente à informação/infra-estrutura que se pretende proteger. Por exemplo, no aeroporto Narita (Tóquio) pretende-se implementar um processo de autenticação que inclui, em conjunção, o reconhecimento de rosto e o reconhecimento da íris. A Central Intelligence Agency (CIA), o Federal Bureau of Investigation (FBI) e a National Aeronautics and Space Administration (NASA) utilizam leitores de retina para proteger o acesso a zonas sensíveis. No entanto, seria excessivamente dispendioso e desajustado utilizar leitores de retina ou de íris para autenticar/identificar o utilizador de um computador pessoal no laboratório pedagógico de informática de uma universidade.

Perceber os níveis de precisão das tecnologias biométricas é uma tarefa difícil, não só pela complexidade dos testes necessários para os conhecer, mas pela dificuldade de obter esses dados do universo de empresas fabricantes destes dispositivos de autenticação. No entanto, é de presumir que as empresas dispostas a fornecer esses dados e/ou sujeitar-se a teste governamentais – como é o caso do "Facial Recognition Vendor Test" do Counterdrug Technology Development Program Office do Departamento de Defesa dos Estados Unidos da América – serão aquelas que se

encontram nos níveis mais avançados de precisão, uma vez que estes testes representam uma forma excelente de publicidade.

Com o objectivo de efectuar uma comparação entre as diversas tecnologias, optou-se por analisar comparativamente produtos do mesmo tipo, seleccionar em cada grupo o(s) mais preciso(s) e, por fim, comparar as várias classes de tecnologias biométricas.

3.2 – Reconhecimento facial

O processo de reconhecimento facial tem início com a captura de uma imagem, seguida da detecção de um rosto que será comparado com modelos armazenados numa base de dados, podendo ser complementada com a análise da cor da pele, detecção de linhas ou ainda de um modelo híbrido [Thian 2001]. As maiores dificuldades neste processo são essencialmente provocadas por diferentes orientações da cabeça [Poh et al. 2001].

Os processos baseados neste tipo de biometria são limitados pelo facto de o utilizador ter que ser enquadrado com o modelo, dada a dificuldade (processamento necessário) em adaptar o modelo à sua cara, isto para além da necessidade de adaptar o modelo a todas as condições que podem alterar a aparência de um indivíduo, como o uso de óculos, envelhecimento, barba, etc. Este processo baseia-se essencialmente na localização de pontos fixos como os olhos, nariz e boca [Poh et al. 2001][Thian 2001].

Os casinos têm utilizado esta tecnologia com sucesso para criar uma base de dados de faces de burlões, de modo a facilmente serem identificados pela segurança [Liu et al. 2001].

O *Counterdrug Technology development Program Office* do Departamento de Defesa dos Estados Unidos da América promoveu o *Facial Recognition Vendor Test 2002* (FRVT2002) num esforço internacional de colaboração com diversas entidades governamentais como, por exemplo, o FBI, o *Canadian Passport Office*, o *Australian Customs* e o *United Kingdom Biometric Work Group*. Participaram dez algoritmos neste teste de grupo. Os valores exactos de FAR e FRR não se encontram disponíveis. No entanto, por observação dos gráficos, podemos obter valores aproximados. A tabela 1 sistematiza os desempenhos do melhor algoritmo, considerando diversas combinações dos dois parâmetros, FAR e FRR.

FAR	FRR%	
0,0001	27,5	**Masculino**
	29	**Feminino**
0,001	9	**Masculino**
	12	**Feminino**
0,01	9	**Masculino**
	11	**Feminino**
0,1	4	**Masculino**
	5	**Feminino**

Tabela 1 – Precisão do Reconhecimento facial.

Da informação disponibilizada, pode-se concluir que a autenticação e/ou identificação com recurso a esta classe de tecnologias é mais precisa em indivíduos do sexo masculino do que em indivíduos do sexo feminino. Ainda assim, a maturidade actual destas tecnologias parece estar ao nível em que se encontravam em 1998 as tecnologias biométricas baseadas na impressão digital [Phillips, 2003], com a agravante de ser uma biometria muito dependente das condições de ambiente como se pode observar, na figura 10, pela variação da taxa de verificações (igual a 1-FRR) em contextos diferentes.

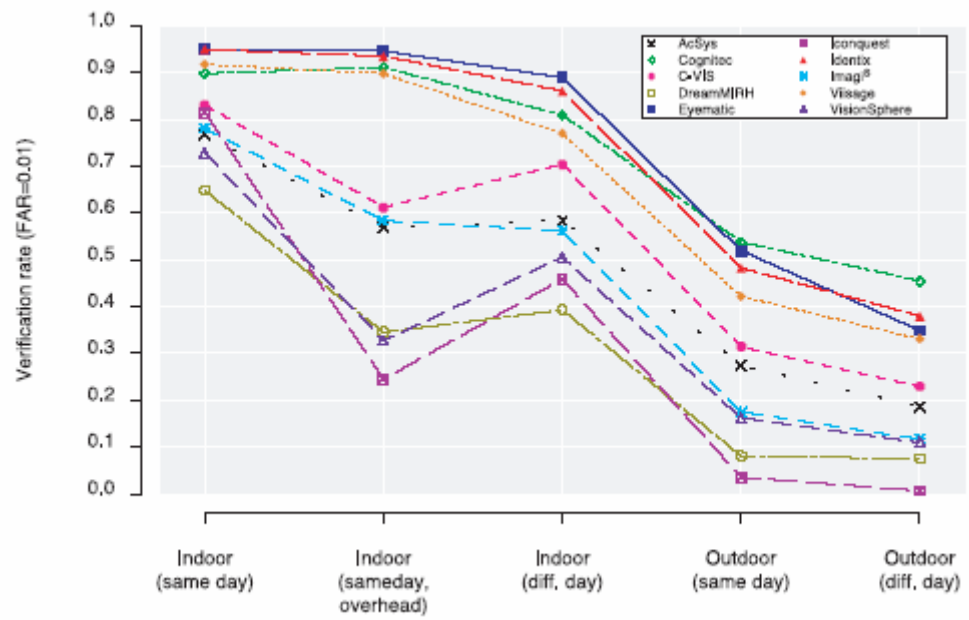


Figura 10 - Performance de algoritmos de reconhecimento facial em diferentes contextos. Fonte: [Phillips, 2003]

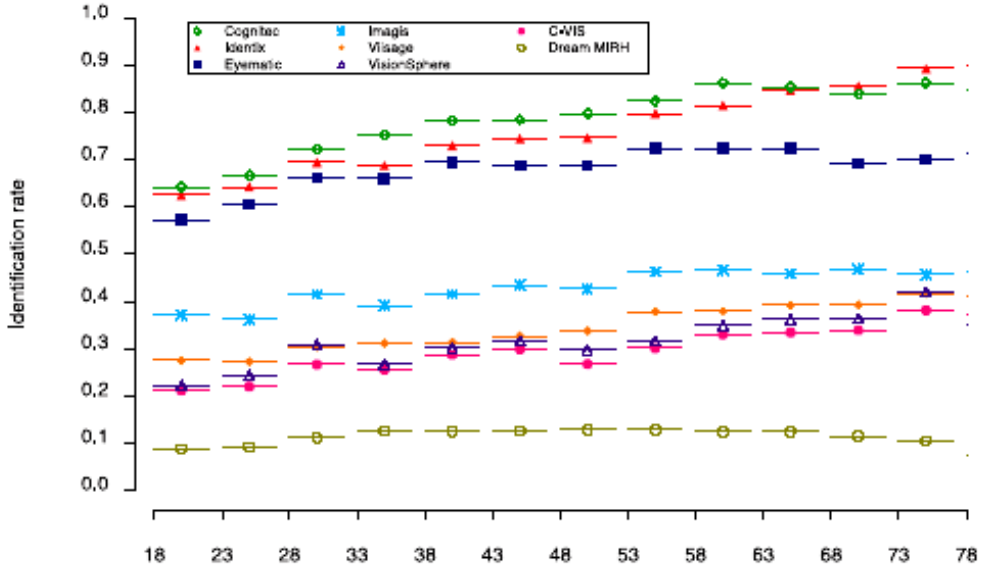


Figura 11 – Performance de algoritmos de reconhecimento facial de acordo com a idade do utilizador. Fonte: [Phillips, 2003]

Outra característica interessante dos algoritmos de reconhecimento facial que se apresentaram ao FRVT2002 é a sua dependência da idade do utilizador. Os algoritmos são, tendencialmente, mais precisos ao avaliar indivíduos mais velhos (figura 11).

3.3 – Geometria da mão

O reconhecimento da geometria da mão resulta de uma análise das características da mão como a forma, o comprimento dos dedos e as suas linhas características. Podemos ter diferentes níveis de segurança neste sistema consoante se utilizem as características em si, a posição das características relativamente a um ponto fixo ou a fixação de vários pontos e as distâncias das características relativamente a todos eles.

De realçar que a geometria da mão (tal como os algoritmos de hoje a interpretam) não é uma característica própria de cada indivíduo, mas tem a vantagem de

facilmente ser combinada com outras biometrias como, por exemplo, a impressão digital [Ross, 1999].

Por outro lado, a geometria da mão, comparada com outras biometrias, não produz um grande conjunto de dados. Portanto, dado um grande número de registos, a geometria da mão pode não ser capaz de distinguir um indivíduo de outro com características da mão semelhantes [Thian 2001].

São poucos os dados técnicos relativos a sistemas biométricos baseados na geometria da mão. No entanto, foi recentemente apresentado um sistema misto que divulgou os valores de FRR e FAR obtidos para o sistema de geometria da mão isoladamente. A FRR obtida foi de 8,34%, enquanto que a FAR foi de 5,29% [Kumar et al, 2003].

Estes valores mostram que, isoladamente, esta tecnologia está longe da maturidade. Quando combinada com outros factores inerentes à mão, como as linhas da palma, os valores melhoram consideravelmente, não pela precisão da representação mas por factores ligados aos algoritmos de decisão.

3.4 – Impressão digital

É, sem dúvida, a tecnologia biométrica mais utilizada actualmente. Embora a generalidade dos sistemas utilizados tenham um nível de fiabilidade muito baixo, esta biometria tem um nível de aceitação muito satisfatório, provavelmente devido ao facto de a impressão digital ser há muito tempo utilizada nos registos civis, associada a documentos de autenticação e/ou identificação.

Os equipamentos normalmente utilizados para a captura dos padrões não distinguem, eficientemente, um dedo vivo de um dedo morto (separado do utilizador legítimo ou replicado sinteticamente) e é muito fácil produzir uma impressão digital sintética com ou sem a colaboração do seu proprietário. Os passos necessários para criar uma impressão digital sem colaboração do seu proprietário são descritos em [Putte et al., 2000]:

- Obter um objecto, como por exemplo um copo, onde o proprietário tenha deixado a sua impressão digital.
- Espalhar delicadamente qualquer tipo de pó fino sobre a zona onde se encontra a impressão utilizando um pincel.
- Colar uma banda de fita cola (fina e transparente) sobre o pó e remove-la.

- Colar a fita-cola no lado fotossensível de um negativo fotográfico e fotografar uma fonte de luz difusa.
- Depois de revelado, o negativo é colocado sobre uma placa fotossensível (como as usadas nos circuitos impressos) e exposto a luz ultravioleta. Retira-se então o negativo.
- Utilizando um banho de gravura com água-forte, as partes da placa expostas à luz ultravioleta são removidas.
- Um último banho de água-forte cauteriza a camada de cobre resultando num perfil muito fino (cerca de 35μ) que é uma cópia "exacta" da impressão original.
- Após aprofundar as marcas de modo a assemelhar-se a uma impressão digital pode ser feito um carimbo de cimento de silicone à prova de água para substituir a impressão digital original.

Existem leitores que tentam ultrapassar o "efeito dedo morto" recorrendo a sensores de tensão arterial, condutividade, temperatura e leitura de padrões existentes em camadas inferiores à epiderme. No entanto, estas tecnologias são caras e ainda não atingiram o nível de maturidade desejado.

O FVC (*Fingerprint Verification Competition*) é um teste de grupo organizado pela Universidade de Bolonha, pela Universidade Estadual San Jose e pela Universidade Estadual do Michigan desde 2000. Enquanto o FVC 2000 contou com a participação a concurso de 11 algoritmos [Maio, 2001], o FVC2002 contou com já com a participação de trinta e um algoritmos, entre produtos académicos, industriais e anónimos [Maltoni et al, 2003]. Da informação disponibilizada por [Maio et al, 2002] e [Maltoni et al, 2003] pode-se concluir (tabela 2) da existência de um grau de maturidade já bastante elevado. No entanto, existem sistemas (mesmo comerciais) muito distantes dos valores desejados, com taxas médias de intersecção de erros (médias porque resultam do cálculo da média das taxas de intersecção de erros obtidas pelo algoritmo nas oito bases de dados utilizadas) superiores a 5%, isto é, taxas vinte e cinco vezes mais altas do que o produto melhor classificado. Devido ao grande número de métricas utilizadas, que dificultam a percepção da precisão dos sistemas avaliados, foram consideradas apenas as dez melhores taxas de intersecção de erros.

Produto	EER (%)
Bioscrypt Inc.	0,19
Anónimo	0,33
Anónimo	0,41
Bioscrypt Inc.	0,77
Siemens AG	0,92
Neurotechnologija Ltd.	0,99
SAGEM	1,18
Andrey Nikiforov (independente)	1,31
SAGEM	1,42
Deng Guoqiang (independente)	2,18

Tabela 2 – Precisão do reconhecimento da impressão digital.

Os leitores de impressão digital vêm, muitas vezes, incorporados em hardware de utilização comum como, por exemplo, o teclado. Torna-se então necessário conhecer o nível de precisão destes dispositivos. A única empresa que respondeu a esta questão indicou que a FAR é menor que 1% e a FRR é inferior a 2%. Aliás, a qualidade da impressão digital capturada pode variar imenso. Na figura 12 podemos observar nove impressões digitais (de dedos diferentes), ordenadas pela sua qualidade.

Figura 12 – impressões digitais com qualidade diferente. Fonte: [maio, 2001]

No FVC2004 obtiverem-se taxas médias de intersecção de erros com valores mais altos do que no FVC2002, denotando que o conjunto de dados biométricos agora em teste é mais exigente para os algoritmos do que o anterior. Todos os dados apresentados [Maio, 2004] na tabela 3 referem-se à competição aberta, isto é, com limites mais permissivos para o tempo de execução. Existem outros resultados referentes à competição "*light*" onde são exigidos tempos de processamento mais rápidos mas que, naturalmente, apresentam resultados, do ponto de vista da fiabilidade, mais fracos. Embora [Maio et al, 2004] divulgue apenas os códigos de referência dos algoritmos é possível identificar os seus criadores na página oficial do evento em http://bias.csr.unibo.it/fvc2004.

Produto	**EER (%)**	**FRR(%) para obtenção de uma FAR nula**
Byoscript Inc. (Canadá)	2,07	6,21
Sonda, Ltd (Rússia)	2,10	6,59
Institute of Automation, The Chinese Academy of Sciences (China)	2,30	10,01
Gevarius (Rússia)	2,45	7,34
Jan Lunter (França)	2,90	32,13

Tabela 3 – Fiabilidade do reconhecimento da impressão digital segundo o FVC2004

3.5 – Leitura de Íris

Esta tecnologia envolve a análise do anel colorido que cerca a pupila do olho humano e é considerada por muitos a menos intrusiva de todas, funcionando mesmo com óculos postos [Liu et al. 2001].

A leitura de íris fornece padrões de comparação com eficácia acima da média e é uma das poucas tecnologias biométricas que pode ser adequada para identificação. No entanto, a dificuldade de utilização e integração com os sistemas existentes é um obstáculo à sua utilização [Liu et al. 2001].

O baixo custo do equipamento necessário é uma vantagem, já que uma câmara normal pode ser utilizada no processo. No entanto, a qualidade da imagem a utilizar no processo é uma questão importante a ter em conta [Thian 2001].

Esta tecnologia é considerada como uma das tecnologias biométricas mais precisas. Wang [Wang 2003] apresenta um trabalho em que combina esta tecnologia com o reconhecimento facial e indica, entre outros, os valores de FRR e FAR para o reconhecimento da íris apresentados na tabela 4.

FAR	FRR
0	0,002
0,2	0,0014
0,6	0,0008

Tabela 4 – Fiabilidade do reconhecimento por leitura da íris

3.6 – Leitura de retina

Os sistemas biométricos baseados na leitura de retina analisam a camada de vasos sanguíneos situada na parte de trás do olho, através da utilização de uma fonte de luz de baixa intensidade, para opticamente reconhecer padrões únicos. Esta tecnologia pode atingir altos níveis de precisão, mas requer que o utilizador olhe para dentro de um receptáculo e foque um determinado ponto, o que não é conveniente para utilizadores que usem óculos ou que receiem o contacto próximo com o leitor [Liu et al. 2001].

O custo do equipamento necessário para a implementação desta tecnologia era, até há muito pouco tempo, um factor limitativo, uma vez que os sistemas biométricos de leitura de retina implementados eram soluções proprietárias desenvolvidas especificamente para as entidades que as exigem. Assim, não existiam no mercado soluções *pret-a-porter*. No entanto, as patentes norte-americanas número 5673097 e 6453057 apresentam soluções de alta portabilidade que poderão revolucionar este mercado. A *Retinal Technologies, LLC* anunciou recentemente que se prepara para colocar no mercado leitores de padrões de retina a um custo extremamente baixo e de alta precisão. Quando questionada quanto aos valores de FRR e FAR, a empresa apresentou um relatório técnico de onde se extraiu a tabela 5, que sistematiza os valores de FRR e FAR em função do valor definido de tolerância. Os valores apresentados representam um nível de precisão incomparável com qualquer outra tecnologia

biométrica. Infelizmente, não foi possível obter informações relativas à precisão ou ao preço de outros sistemas anunciados.

Tolerância (t)	FAR	FRR
0.37	2.38E-5	0.000271
0.41	3.38E-6	0.00081
0.44	4.17E-7	0.0022
0.48	4.48E-8	0.0055

Tabela 5 – Fiabilidade anunciada de um sistema leitor de retina

3.7 – Análise comparativa (por classes) da precisão das tecnologias biométricas físicas

De modo a facilitar a comparação, a tabela 6 e o gráfico da figura 13 sintetizam os valores encontrados para FRR e FAR das várias técnicas acima descritas. Para a impressão digital considerou-se, para o melhor algoritmo encontrado, o valor de CER e o valor de FRR que anula a FAR.

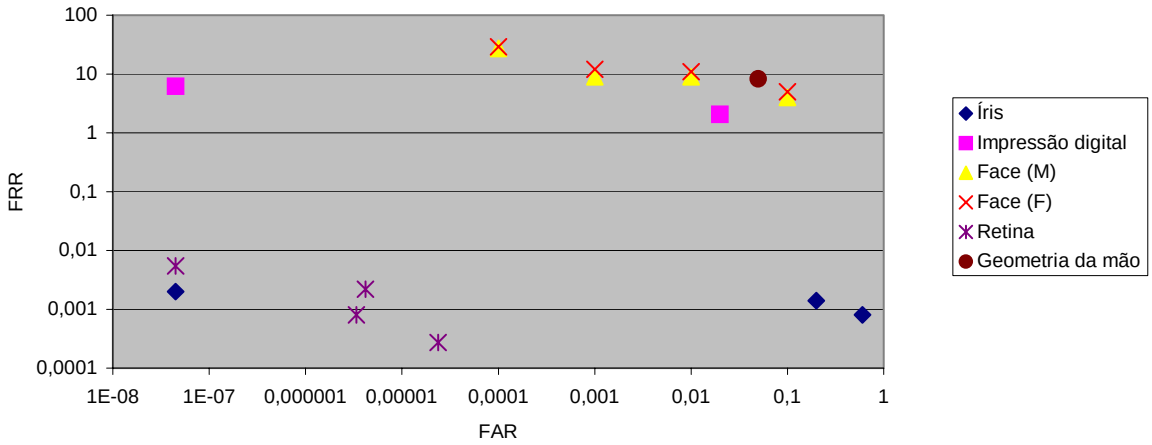


Figura 13 – FRR vs FAR das várias tecnologias (escala logarítmica)

No gráfico da figura 13 utilizam-se eixos logarítmicos para uma melhor observação da posição relativa das diferentes técnicas. Como não é possível representar o valor zero neste tipo de gráfico e como o zero referido na tabela 6 para a FAR é um valor teórico cujo significado não é mais do que "quase nulo", uma vez que o nível de precisão é dependente do número de dados da amostra de teste e esta nunca coincide com a população, optou-se por representar os respectivos valores de FRR no ponto (4,48E-8, FRR).

FAR	Íris	Impressão digital	Face (M)	Face (F)	Retina	Geometria da mão
4,5E-08	0,002	6,21			0,0055	
4,2E-06					0,0022	
3,4E-06					0,00081	
2,4E-05					0,000271	
0,0001			27,5	29		
0,001			9	12		
0,01			9	11		
0,02		2,07				
0,05						8,34
0,1			4	5		
0,19						
0,2	0,0014					
0,33						
0,41						
0,6	0,0008					
0,77						
0,92						
0,99						

Tabela 6 – FRR vs FAR das várias tecnologias estudadas.

3.8 – Reconhecimento de voz

Os processos de autenticação que recorrem ao reconhecimento da voz baseiam-se no facto de as características físicas de cada indivíduo, associadas a hábitos comportamentais, proporcionarem à sua voz características únicas que podem ser representadas por um espectro de frequências. A figura 14 mostra o espectro amplitude

vs frequência da voz de Alanis Morissette construído com o Xanalyser, uma ferramenta de análise de frequência para o Xwindow. Pode-se observar que existe um pico a cada 500Hz. No entanto, a informação capturável não possui informações suficientes para garantir o reconhecimento (identificação) em larga escala de indivíduos [Jain et al. 2000].

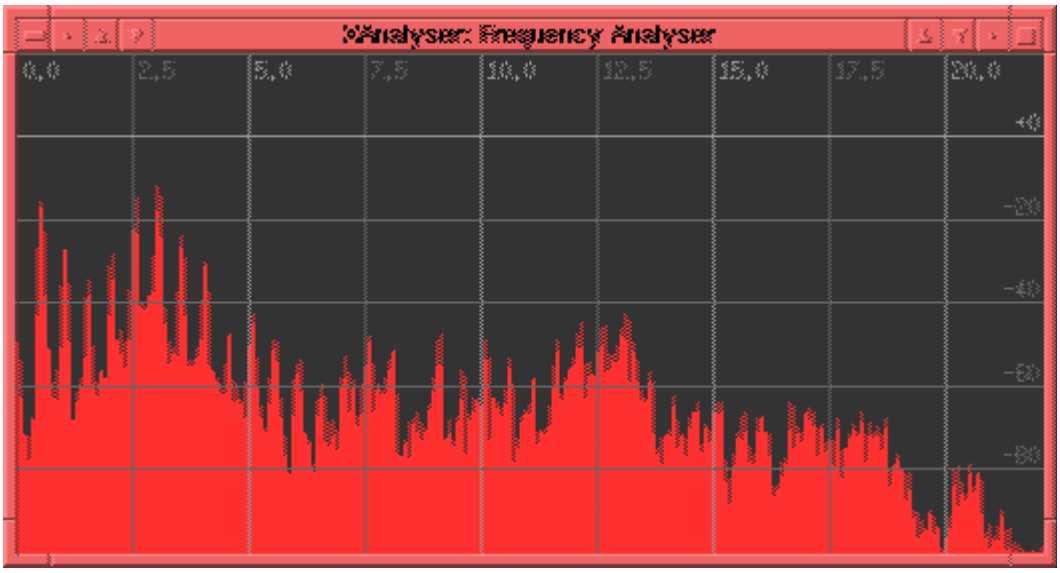


Figura 14– Gráfico amplitude vs frequência da voz de Alanis Morissette.
Fonte: http://www.suse.de/~arvin/xanalyser

Estes processos fundamentam-se nas técnicas de processamento de voz e na biometria, e o envolvimento do utilizador pode passar pela introdução (oralmente) no sistema de uma palavra/frase chave ou pela leitura de um conjunto de caracteres que, combinados, fornecem um conjunto de características suficientes para permitir a autenticação ou a identificação do indivíduo. [Markowitz, 2000]

O potencial destes sistemas é grande devido ao baixo custo do hardware necessário que, aliás, está já presente em grande parte dos computadores existentes: um microfone. No entanto, a sua aplicação está limitada, actualmente, a aplicações com um baixo nível de segurança, em virtude das grandes variações na voz de um indivíduo e na baixa precisão dos actuais sistemas de autenticação por reconhecimento de voz.

3.9 – Assinatura manual recolhida de modo digital

A assinatura tem sido utilizada como um elemento de identificação largamente disseminado. É utilizada para comprometer indivíduos e organizações em contratos e para realizar pagamentos através de, por exemplo, cartões de crédito. A assinatura manual pode ser utilizada como uma biometria para autenticação e/ou identificação desde que se possua um painel que capture a velocidade e a pressão dos movimentos que geram a assinatura, bem como a sua forma.

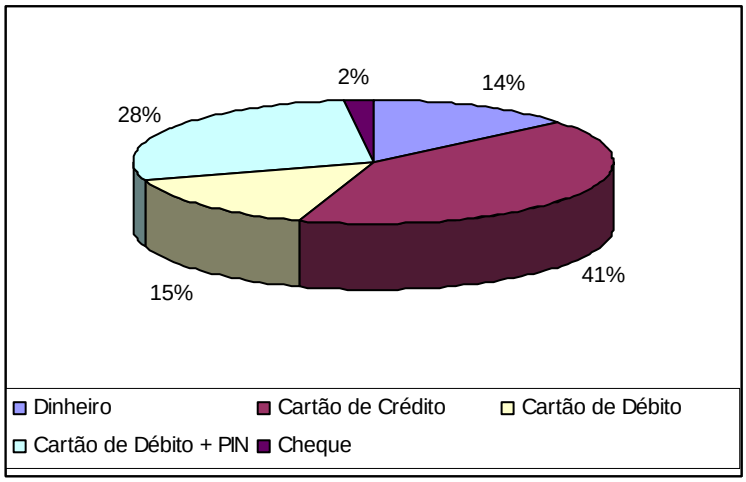


Figura 15 – Resultado do inquérito realizado pela epaynews em Janeiro de 2004 [Epaynews, 2004] com a questão *"As a consumer, which of the following payment methods are you most comfortable with?"* (como consumidor, com qual dos seguintes métodos de pagamento se sente mais confortável?)

Segundo um inquérito realizado pela Epaynews, os cartões de crédito e os cheques são os meios de pagamento preferidos por 43% dos utilizadores (figura 15) e, uma vez que são utilizadas em conjunto com a assinatura do seu proprietário, podem tornar esta tecnologia numa das mais disseminadas, atenuando as questões relacionadas com a sua utilização fraudulenta.

3.10 – Dinâmica de digitação

A técnica habitualmente designada por *Keystrokes Dynamics*, também conhecida nos países de língua oficial portuguesa por dinâmica de digitação, é baseada na monitorização dos padrões comportamentais do utilizador ao digitar palavras/frases

passe e/ou texto durante uma sessão. Regra geral, o sistema requer que o utilizador, na primeira utilização, digite a mesma frase um determinado número de vezes. Contudo, teoricamente um sistema pode, na primeira utilização, recolher a informação necessária para encontrar um padrão, sem o conhecimento do utilizador. É também possível ao sistema adaptar o modelo do padrão ao longo do tempo, de forma a ajustar-se a novos padrões recolhidos.

No que concerne à verificação de padrões, além dos métodos que recorrem a redes neuronais, foram desenvolvidos e testados diversos algoritmos baseados na estatística e nas probabilidades. As descrições que se seguem não são matematicamente rigorosas uma vez que apenas pretendem apenas apresentar os princípios básicos dos algoritmos determinísticos, com melhor desempenho.

Joyce [Joyce, 1990] apresenta um método baseado no cálculo de um valor representativo da distância dos tempos de latência entre os caracteres introduzidos e os tempos de latência previamente armazenados. Este algoritmo começa por calcular, para o conjunto de tempos de latência medidos, a média e o desvio padrão (oito entradas do nome de utilizador, palavra passe, primeiro e último nome) e, de seguida elimina os que estão fora de 3 desvios padrão. O processo é repetido de forma a criar uma *assinatura* do utilizador. No processo de autenticação calcula $\| M - S \| = \sum_{i=1}^{8} \sqrt{(M_i - S_i)^2}$, em que M_i é a média de cada uma das 8 entradas e S_i é o valor de cada entrada. Calcula a média e o desvio padrão dos 8 módulos e define o intervalo de decisão em Média±1,5*DesvPad. O utilizador é aceite se, pelo menos, 60% dos tempos de latência estiverem dentro dos respectivos intervalos de decisão.

De modo a apresentar um método de identificação baseado na dinâmica de digitação [Monrose, 1997] recorre à distância euclidiana e a cálculos probabilísticos baseados na presunção de que os tempos de latência para uma mesma sequência seguem uma distribuição normal. Neste algoritmo, os utilizadores registados estão agrupados em classes de acordo com o número de palavras digitadas por minuto e possuem uma assinatura digital construída da mesma forma que em [Joyce, 1990]. Embora este algoritmo implique que sejam digitados um elevado número de palavras, ele pode ser adaptado para uso com frases passe se o critério de construção das classes for o número de caracteres por segundo. O processo de identificação começa pela escolha da classe em que os dados, um vector de dimensão N ainda anónimo, serão incluídos para

comparação. O vector de tempos recolhido, *U*, é comparado com os vectores *R* da classe a que pertence e das classes mais próximas, de acordo com o número de palavras digitadas por minuto, e é encontrado o vector que maximiza a probabilidade calculada pelo classificador: $Score(R,U) = \sum_{i=1}^{N}\left(\left[\frac{1}{o_{u_i}}\left[\sum_{j=1}^{o_{u_i}} P\left(\frac{X_{ij}^{(u)} - \mu_{r_i}}{\sigma_{r_i}}\right)\right]\right] * weight_{u_i}\right)$, onde :

- *P* é a Função Densidade de Probabilidade Normal;
- σ_{r_i} é o desvio padrão de todos os tempos em *R* correspondentes ao mesmo par de caracteres de onde resulta o valor *i* do vector *U*;
- μ_{r_i} é a média de todos os tempos em *R* correspondentes ao mesmo par de caracteres de onde resulta o valor *i* do vector *U*;
- o_{u_i} é o número de ocorrências em *R* de tempos correspondentes ao mesmo par de caracteres de onde resulta o valor *i* do vector *U*;
- $X_{ij}^{(u)}$ é o valor (tempo) da j-ésima ocorrência de tempos correspondentes ao par de caracteres que originaram o i-ésimo valor do vector *U;* e
- $weight_{u_i} = \begin{cases} 0 & \text{se } u_i \text{ ou } r_i \text{ são vazios} \\ \dfrac{o_{u_i}}{\sum_{k=1}^{N} o_{u_k}} & \text{caso contrário} \end{cases}$

Resumidamente, podemos dizer que o classificador de um determinado valor do vector *U* é a probabilidade, Normal com desvio padrão σ e média μ, de ele ser encontrado no vector *R*.

Os autores de [Monrose, 1997] colocaram ainda a hipótese dos métodos de autenticação/identificação por *Keystroke Dynamics* permitirem quebrar a segurança das chaves geradas pelo software de segurança PGP – Pretty Good Privacy – por este utilizar os tempos de digitação como fonte de aleatoriedade das sementes geradoras utilizadas nos algoritmos de encriptação e assinatura de mensagens. Posteriormente, em [Monrose, 2001] descreve-se um algoritmo que recorre à Álgebra, nomeadamente aos polinómios e aos espaços vectoriais, para gerar palavras passe complexas partindo de uma palavra passe simples e do padrão de digitação do utilizador.

Monrose [Monrose, 2000] apresenta um sistema de identificação baseado nos modelos de semelhança de Bayes, com o objectivo de permitir a identificação de um utilizador. Para tal, os utilizadores registados são agrupados de acordo com os grupos de caracteres em que têm ritmos de digitação semelhantes. Assim, por exemplo, os elementos do grupo correspondente ao conjunto {as, ta, mo, de} têm ritmos semelhantes ao digitar estas 4 sílabas, mas têm ritmos distintos (nestas sílabas) dos membros de todos os outros grupos. Quando um conjunto de dados é recolhido para identificação do seu utilizador, é constituído para cada grupo um vector *x* de dimensão *n* com os tempos correspondentes às suas *n* sílabas. Presumindo que estes vectores seguem uma distribuição Normal, o vector desconhecido é associado à pessoa que tem uma probabilidade maior de ser o seu "proprietário". A função de decisão, Δ, para o cálculo da distância entre dois vectores *x* e *x'* é:

$$\Delta^{\alpha}(x, x') = \sum_{i=1}^{n} w_i \left(\frac{\sqrt{(x_i - x'_i)^2}}{\sigma_i} \right)^{\alpha}$$

onde w_i é uma função ponderadora calculada dividindo a frequência absoluta da sílaba *i* (em todos os dados de todos os grupos), pela frequência absoluta de todas as sílabas (em todos os dados de todos os grupos); σ_i é o desvio padrão dos tempos correspondentes à sílaba *i* no conjunto dos dados e α é uma constante que serve para ajustar a robustez do algoritmo (valores mais próximos de 1 do que de 2 aumentam ligeiramente a eficácia do algoritmo).

Os algoritmos apresentados são uma pequena amostra das várias aproximações utilizadas para encontrar algoritmos de *Keystroke Dynamics* que forneçam níveis de segurança, na autenticação satisfatórios. Outros podiam ainda ser referidos, todos com métodos de avaliação diferentes, diferente número de utilizadores envolvidos (normalmente muito reduzido), diferente número de caracteres necessários para fazer o registo no sistema e diferente número de caracteres necessários para proceder à autenticação/identificação de um utilizador. Esta disparidade de parâmetros dos algoritmos e da sua avaliação tornam impossível a tarefa de os comparar. Além disso, não existe, neste contexto, um conceito de amostra representativa. O mesmo algoritmo apresenta resultados diferentes quando testado com diferentes grupos de voluntários. A única forma de comparar dois algoritmos é testá-los com o mesmo grupo de dados e concluir qual é o melhor para o nosso caso. No que respeita a aplicações Web, onde este

método não é exequível, devemos considerar apenas os resultados que envolvam um número considerável de voluntários. O esforço computacional do algoritmo deve também ser tido em conta, uma vez que o tempo de execução é um factor crítico em qualquer aplicação.

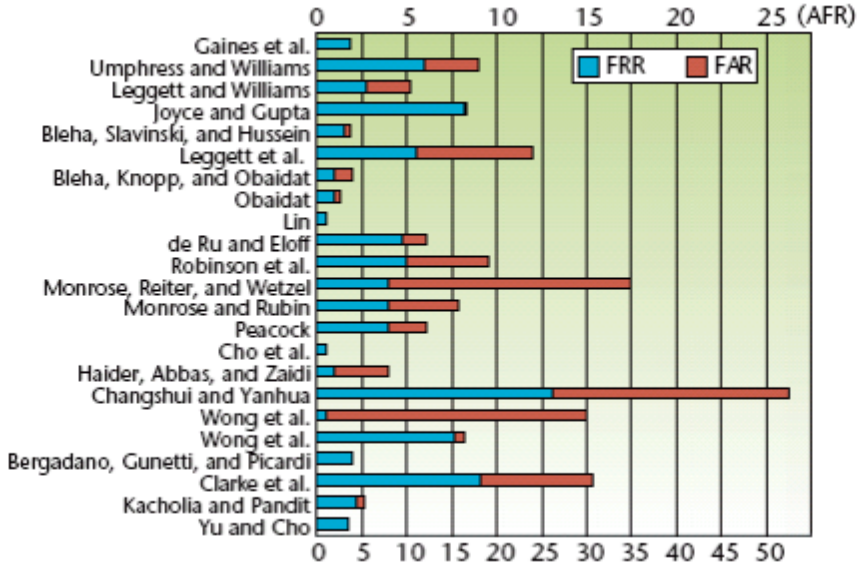


Figura 16 – Precisão dos algoritmos de Keystroke Dynamics. Fonte: [Peacock, 2004]

Ainda assim, é de registar que os valores anunciados da FAR variam desde os 0% até mais de 50%, a FRR varia de mais de 25% a menos de 1% e o número de utilizadores envolvidos na avaliação dos algoritmos está, geralmente, entre dez e cem [Peacock, 2004]. As figuras 16, 17 e 18 foram extraídas de [Peacock, 2004] e sintetizam esses dados. A figura 16 apresenta, para cada algoritmo, a FRR, a FAR e o valor AFR (apresentado no eixo superior) correspondente à média entre a FRR e a FAR; a figura 17 apresenta o número de caracteres necessários para fazer o registo no sistema (a vermelho) e para proceder à autenticação do utilizador (a azul). A figura 18 apresenta o número de utilizadores envolvidos na avaliação de cada algoritmo.

Pode-se observar pelo factos apresentados que não existem algoritmos que apresentem taxas de erro satisfatórias, mantendo uma exigência computacional que lhes permita serem processados dentro de um Smart Card. Foi nesse sentido que se desenvolveu a o trabalho de investigação descrito no capítulo seguinte.

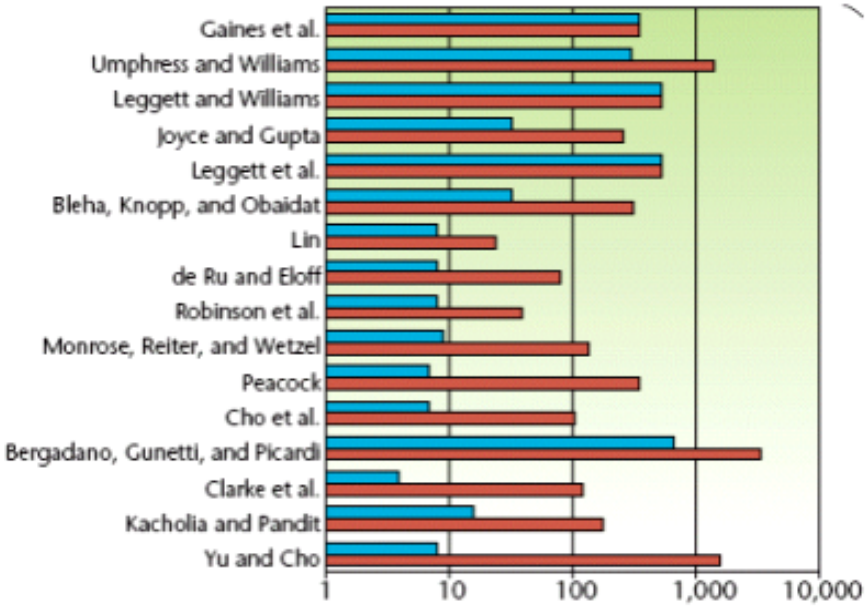


Figura 17 – Número de caracteres necessários. Fonte: [Peacock, 2004]

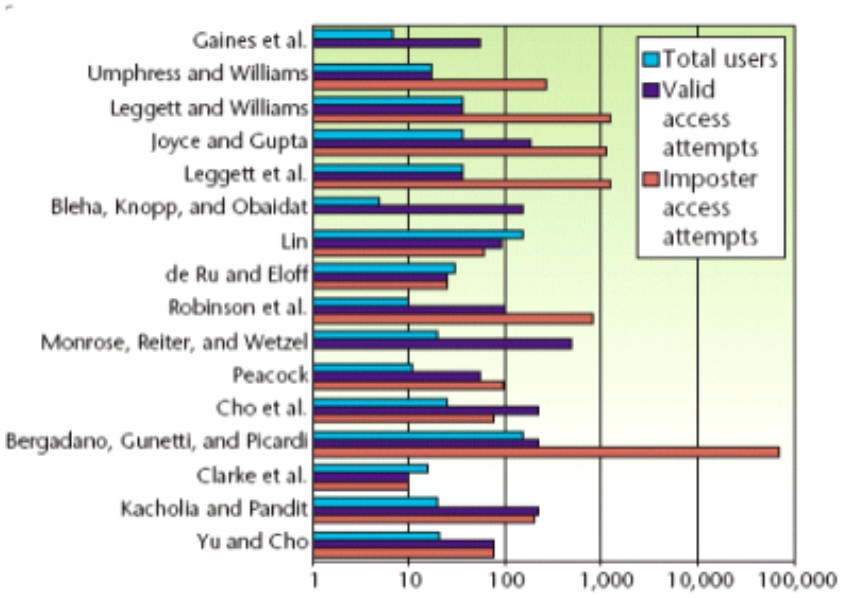


Figura 18 – Número de utilizadores/tentativas dos testes de precisão.
Fonte: [Peacock, 2004]

4 – Um novo algoritmo de Keystroke Dynamics

4.1 – Uma hipótese negada

No início do trabalho de investigação associado a esta dissertação colocou-se a hipótese do padrão de digitação de um indivíduo ser proporcional à distância entre os caracteres digitados e função do número de dedos utilizados para digitar. As figuras 19, 20 e 21 mostram possíveis função distância para utilizadores que digitem, respectivamente, com um dedo, com um dedo de cada mão e com todos os dedos.

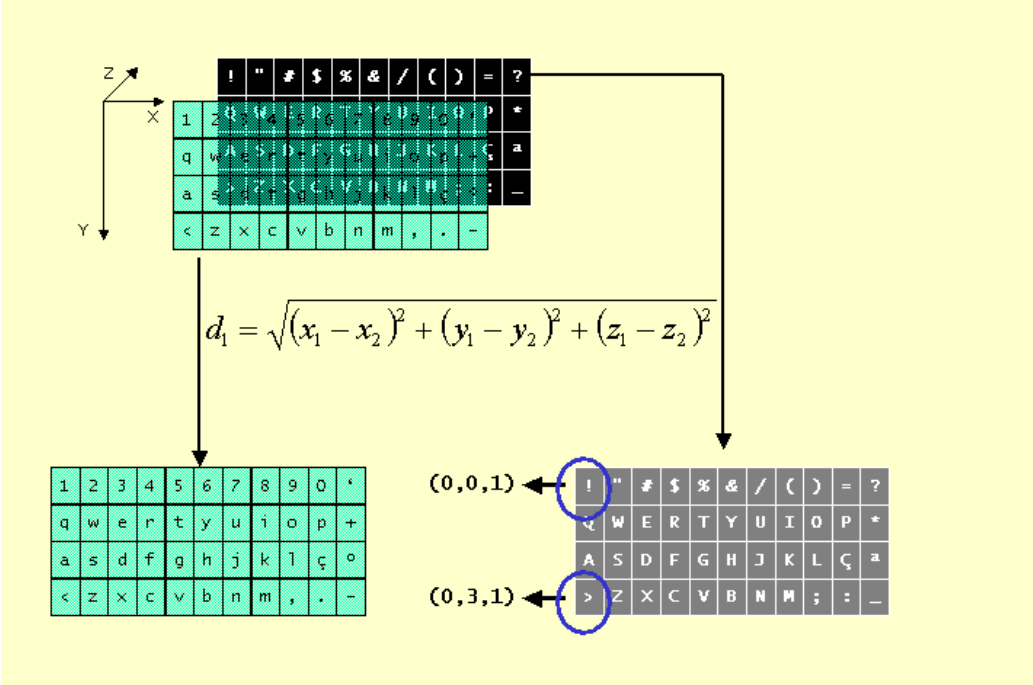


Figura 19– Função distância entre dois caracteres para utilizadores que digitem com um só dedo.

No caso em que o utilizador usa apenas um dedo para digitar (figura 19), todas as teclas de acesso directo no teclado estão num mesmo plano e entre elas aplica-se a distância euclidiana em $\Re^2$. As teclas acessíveis através da tecla SHIFT estão, entre si, na mesma situação. A passagem entre um e outro plano corresponde à passagem de um ponto no plano $z = 0$ para um ponto no plano $z = 1$. Assim, cada tecla corresponde a um vector (x, y, z), com $x \in \{0,1,2,3,...,11\}, y \in \{0,1,2,3\}$ e $z \in \{0,1\}$ e a função

distância, *d*, entre os pontos $a=(a_1,a_2,a_3)$ e $b=(b_1,b_2,b_3)$ é a distância euclidiana, isto é, $d(a,b)=\sqrt{(a_1-b_1)^2+(a_2-b_2)^2+(a_3-b_3)^2}$.

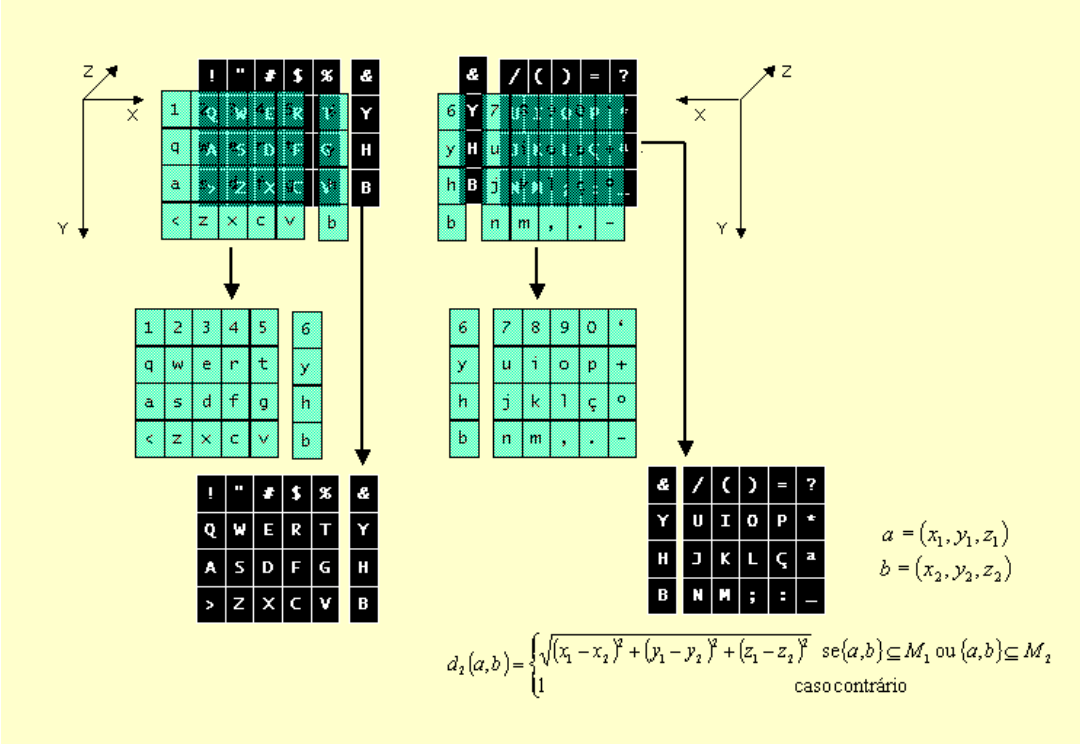


Figura 20– Função distância entre dois caracteres para utilizadores que digitem com um dedo de cada mão

No caso em que o utilizador digita com um dedo de cada mão (figura 20), o teclado fica dividido na zona esquerda e direita. A coluna de caracteres intermédia fica na zona de acesso de ambas as mãos. Uma possível função distância seria aquela que coincide com a distância euclidiana se ambos os caracteres estão na mesma zona do teclado e que atribui zero caso contrário. Esta atribuição do valor zero justifica-se porque, uma vez que são utilizados dedos de mãos diferentes, o dedo que irá digitar o segundo caracter coloca-se em posição durante o movimento da outra mão.

Uma terceira hipótese é o utilizador digitar com todos os dedos (figura 21). Nesse caso as zonas a considerar são as colunas de caracteres. Assim, numa mesma coluna teríamos a distância euclidiana e, entre colunas um valor constante. Este valor teria que ser maior do que zero uma vez que o argumento do pré-posicionamento dos dedos não é aplicável dado o elevado número de caracteres e de dedos envolvidos.

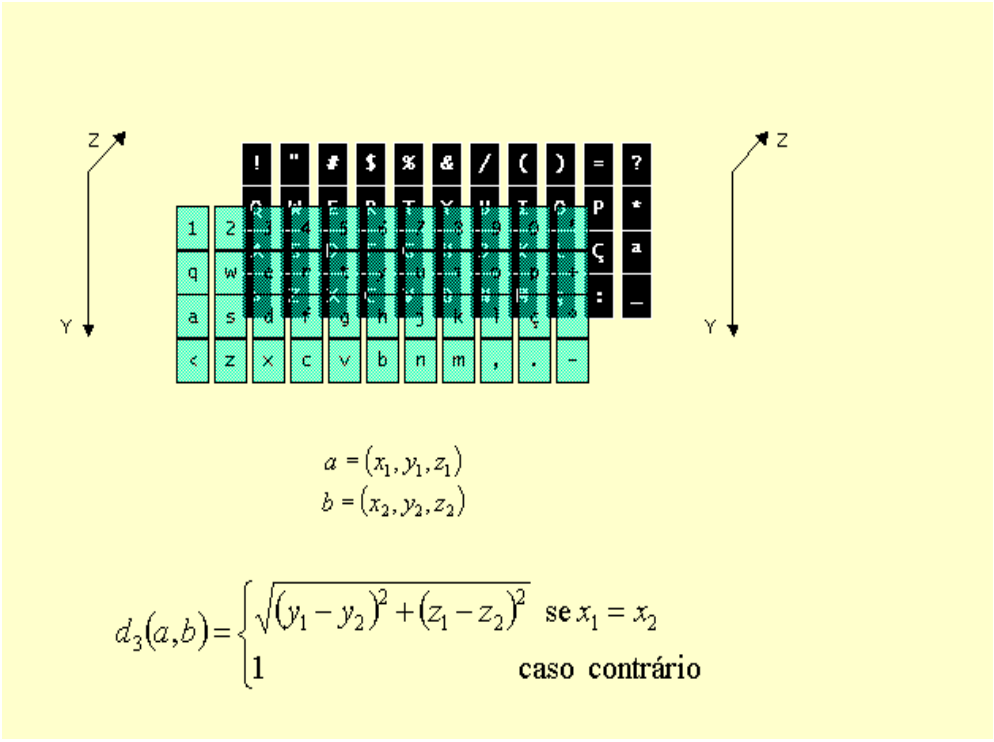


Figura 21– Função distância entre dois caracteres para utilizadores que digitem com todos os dedos.

Todas estas funções distância representaram hipóteses sugeridas pela intuição de que os tempos de latência seriam função da distância dos caracteres e do número de dedos utilizados. A prática demonstrou que não é assim e que a intuição falhou. As figuras 22, 23 e 24 mostram dez tempos de latência para a sequência de caracteres "PO" e os correspondentes tempos de latência da sequência "OP", obtidos a partir da digitação da palavra "POPULAR". Pode-se observar que, embora em certos casos haja semelhanças de comportamento das curvas (figura 22), na generalidade quando o primeiro tempo aumenta o segundo pode aumentar ou diminuir (figuras 23 e 24), contrariando a hipótese colocada uma vez que em qualquer função distância a distância de *a* a *b* é a mesma que de *b* a *a*.

De modo a reforçar a ideia de que, na generalidade, os tempos não são proporcionais a uma qualquer função distância dos caracteres, a figura 25 mostra os tempos de latência entre *P* e *O* e entre *O* e *P* para as primeiras dez tentativas de cada um dos utilizadores do grupo controlado. Infelizmente, a quantidade de dados torna o gráfico menos legível.

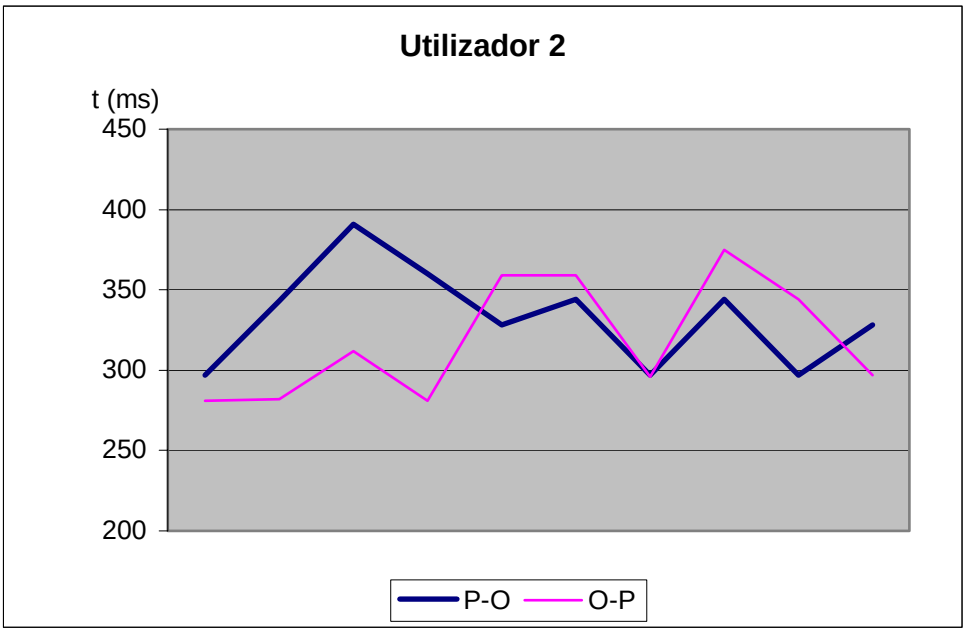


Figura 22 - Tempos de latência do utilizador 2 para as sequências de caracteres PO e OP (dez primeiras tentativas).

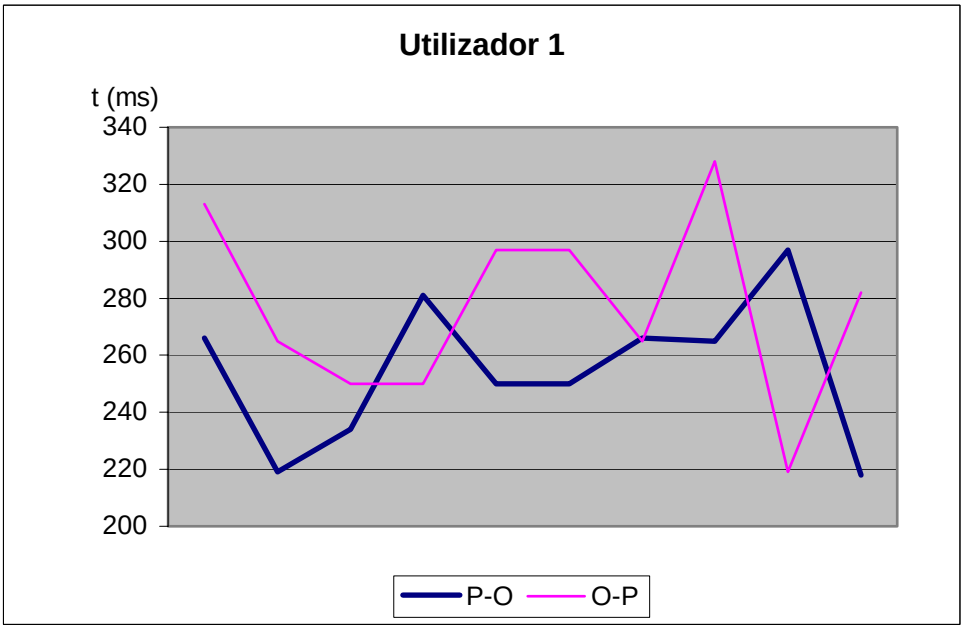


Figura 23 - Tempos de latência do utilizador 1 para as sequências de caracteres PO e OP (dez primeiras tentativas).

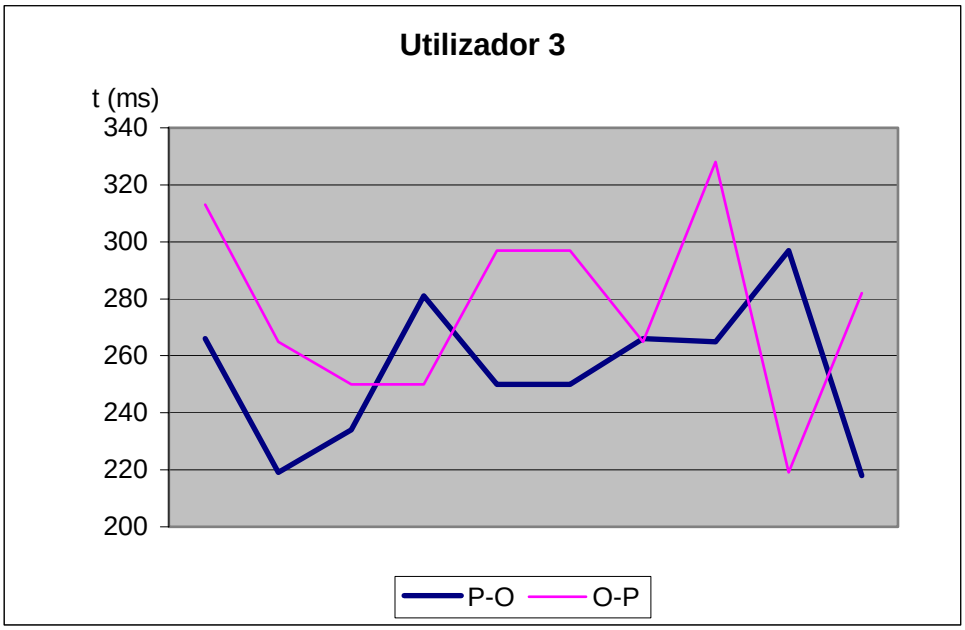


Figura 24 - Tempos de latência do utilizador 3 para as sequências de caracteres PO e OP (dez primeiras tentativas).

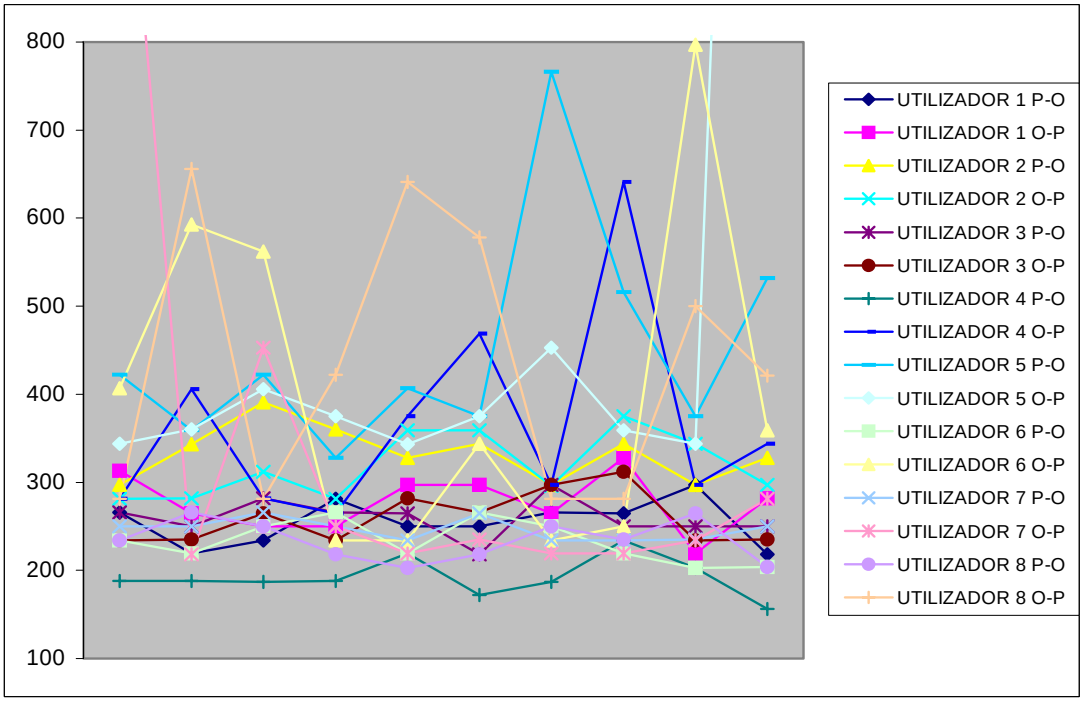


Figura 25 - Tempos de latência, dos utilizadores controlados, para as sequências de caracteres PO e OP (dez primeiras tentativas).

4.2 – Um novo processo de decisão

Dos algoritmos de Keystroke Dynamics apresentados, aqueles que apresentam um baixo nível de exigência computacional de modo a serem suportáveis eficientemente por um servidor Web (autenticação em larga escala) ou por dispositivos de pequena capacidade (como os SmartCards) são os publicados em [Joyce, 1990] e em [Monrose, 1997]. Este último baseia-se, como foi referido, no pressuposto de que os tempos de latência para uma mesma sequência seguem uma distribuição Normal. No entanto, os resultados obtidos durante esta investigação indicam que esse pressuposto está um pouco distante da realidade. A figura 26 mostra uma distribuição tipo dos tempos obtidos e, mesmo com uma regressão polinomial de grau 6, observa-se uma certa distância da distribuição Normal.

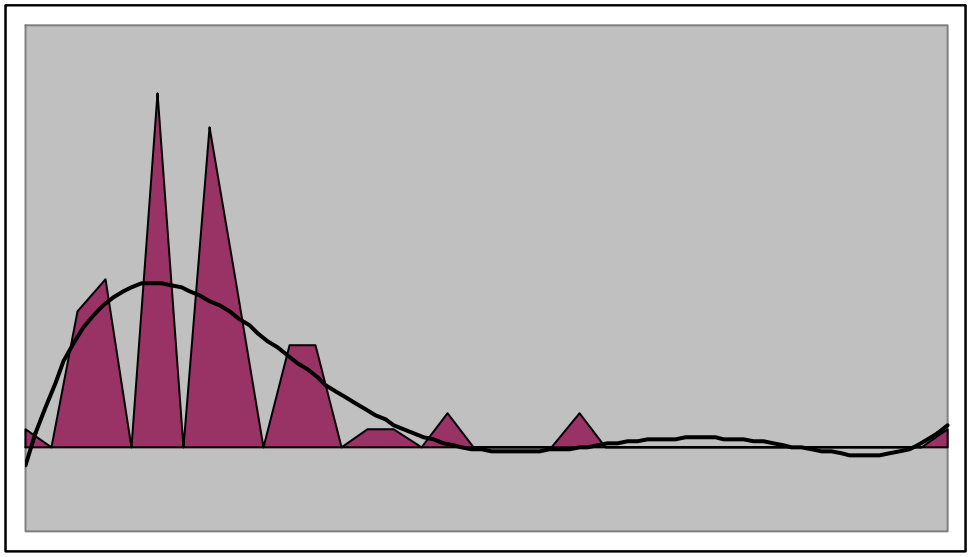


Figura 26 – Uma distribuição dos tempos de latência.

Esta discrepância entre a distribuição observada e a distribuição Normal sugere uma evolução no algoritmo de Joyce & Gupta [Joyce, 1990] que tem no seu algoritmo o cálculo da média, uma medida de tendência central. Uma vez que neste algoritmo é utilizada uma só medida de tendência central, a média, e já que apenas na distribuição Normal a média tem o mesmo valor da moda e da mediana, uma evolução possível é introduzir outra(s) medida(s) de tendência central, como veremos na secção seguinte.

O algoritmo proposto utiliza um conjunto de cálculos pouco exigente do ponto de vista computacional com vista a que um computador, por exemplo um servidor Web, possa autenticar simultaneamente um grande número de utilizadores e a que seja

possível a sua utilização em pequenos dispositivos electrónicos, por exemplo JavaCards. Houve também a preocupação de garantir que o padrão do utilizador pudesse evoluir com o tempo, uma vez que o utilizador pode tornar-se mais rápido à medida que se habitua a digitar a sua palavra/frase passe, assim como pode tornar-se mais lento, à medida que envelhece e fica mais cansado e com menos reflexos. O algoritmo está desenhado para permitir que essa evolução aconteça, lentamente ao longo o tempo, sem que o processo de autenticação seja penalizado.

O processo de registo no sistema é simples: o utilizador digita a sua palavra/frase passe habitual doze vezes. Os dados são gravados e é calculada e gravada a média, a mediana e o desvio padrão dos tempos despendidos entre cada dois caracteres e do tempo despendido na introdução do conjunto de caracteres que forma a palavra/frase passe.

O processo de autenticação é, do ponto de vista do utilizador, igual à autenticação em qualquer sistema que solicite uma palavra/frase passe, tendo apenas que a introduzir, como habitualmente. O sistema irá então comparar os tempos de latência propostos (recolhidos nesta introdução) - TLP - com os armazenados, classificando os primeiros como aceitáveis se e só se

$$Mínimo(média; mediana)*\left(0{,}95-\frac{DesvPadrão}{média}\right) \le TLP \le Máximo(média; mediana)*\left(1{,}05+\frac{DesvPadrão}{média}\right).$$

Cada TLP classificado como aceitável irá contribuir para uma soma, A, com o valor 1 se o TLP anterior não é aceitável (ou se é o primeiro valor) e com 1,5 se o TLP anterior é aceitável. Esta distinção valoriza sequências de caracteres que cumprem os requisitos, desvalorizando desvios pontuais. Inversamente, acertos dispersos são desvalorizados. Os valores não aceitáveis não contribuem para a soma A. O valor final de A decidirá se a identidade do utilizador é, ou não, confirmada, de acordo com o ponto de decisão definido pelo administrador do sistema (a secção 5.2 apresenta a interpretação geométrica da fórmula de decisão).Por exemplo, se o ponto de decisão está definido em 70%, um utilizador só será admitido ao sistema se o valor A da sua palavra/frase passe é superior a 70% do valor máximo possível para A: $(número\_de\_caracteres-1)*1.5+1$. De notar que o número de avaliações é igual ao número de tempos entre caracteres mais um correspondente ao total do tempo decorrido. Finalmente, se e só se a autenticação no sistema é aceite, os tempos de latência mais antigos armazenados são substituídos pelos correspondentes valores

recolhidos nesta tentativa bem sucedida. Este último procedimento permite que os dados armazenados evoluam com o utilizador.

Para que a utilização deste algoritmo disponha de um ambiente seguro de

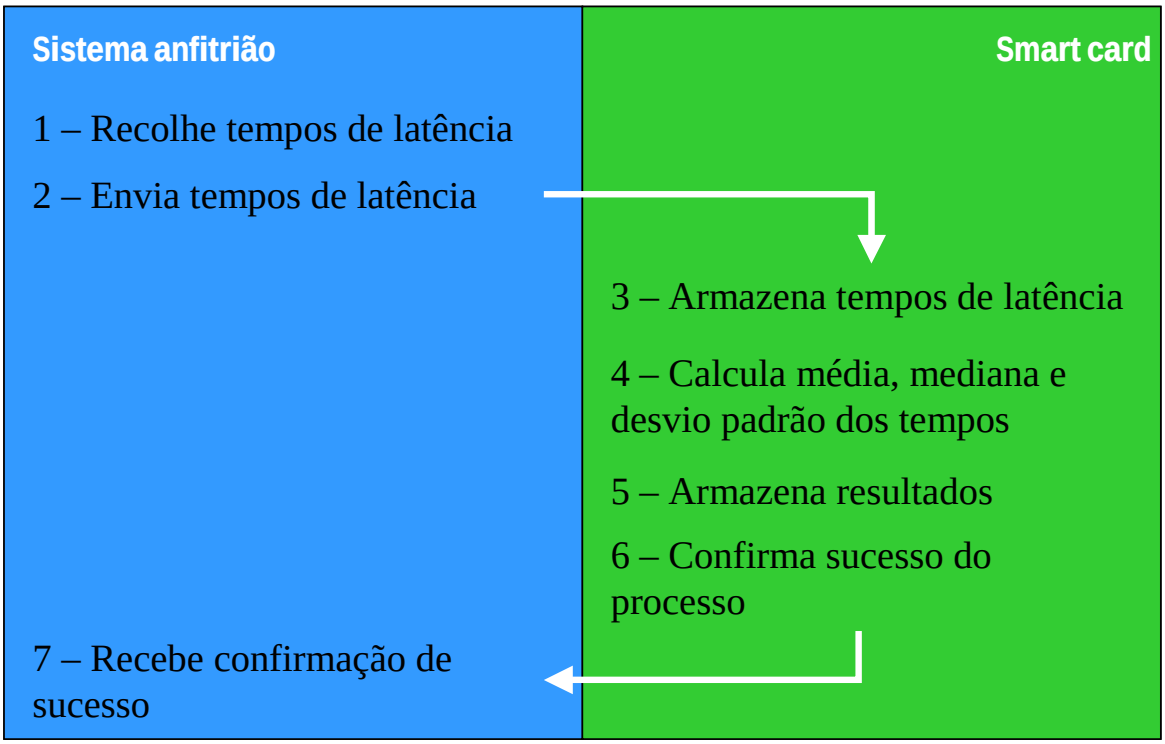


Figura 27 – Distribuição, entre um SmartCard e um CAD, do processo de registo inicial do sistema.

processamento e armazenamento ele deve ser implementado, sempre que o contexto o permita, com recurso a SmartCards. Assim, todos os dados críticos são mantidos e processados dentro do cartão. As figuras 27 e 28 mostram, de modo simplificado, a distribuição do processamento entre o dispositivo anfitrião e o cartão. As comunicações entre o cartão e o sistema anfitrião devem incluir encriptação, apesar do padrão original não sair do cartão, principalmente quando o sistema anfitrião e o dispositivo leitor do SmartCard não se encontram ligados directamente. É de salientar que mesmo em aplicações Web é cada vez mais frequente o uso destes cartões, inclusive como modo de pagamento automático.

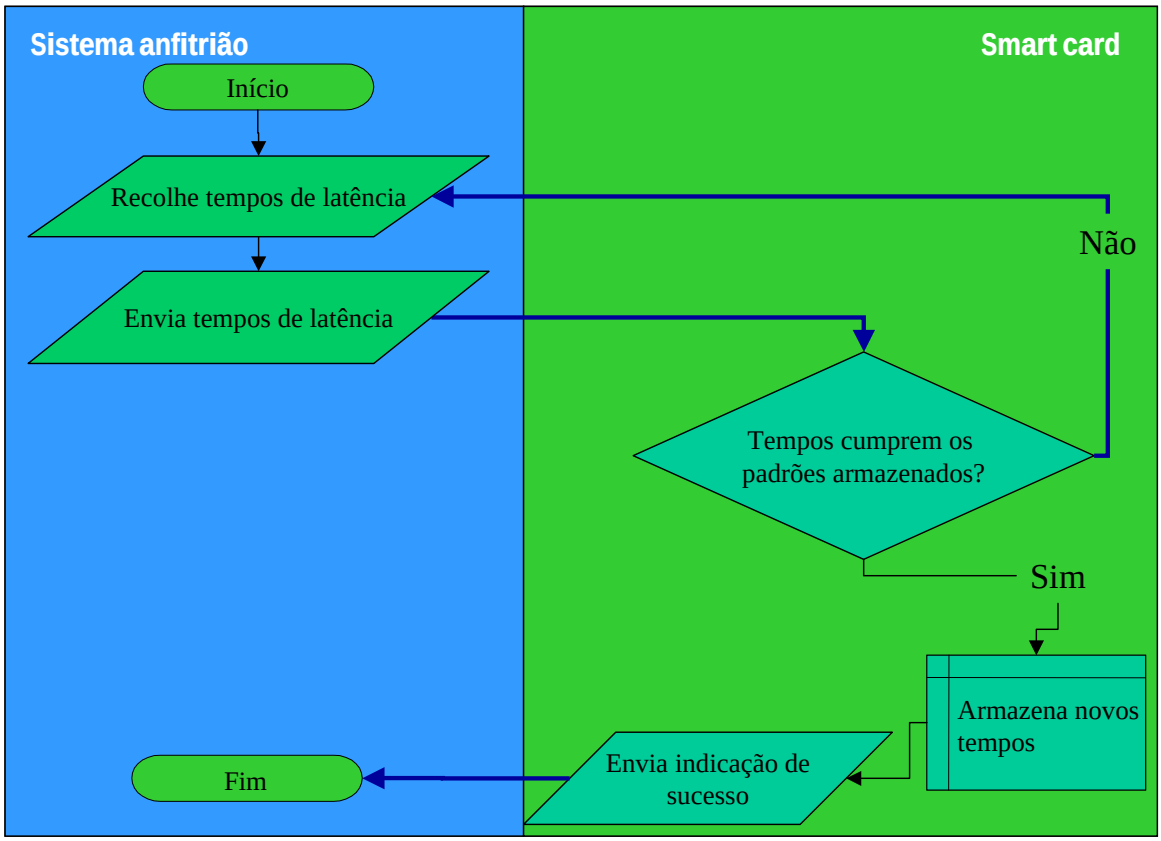


Figura 28 – Fluxograma do processo distribuído de autenticação.

4.3 – Interpretação geométrica da fórmula de decisão

Da presunção de que os tempos de latência seguem uma distribuição Normal, resulta que a moda, a média e a mediana têm o mesmo valor. O algoritmo apresentado não parte dessa premissa (embora assuma uma semelhança entre a distribuição dos dados e a distribuição Normal) e, portanto, inclui o cálculo da média e da mediana. Não foi considerado o cálculo da moda porque não existem dados armazenados em quantidade suficiente para o permitir. As figuras 29 e 30 mostram a relação entre a média e a mediana numa distribuição assimétrica negativa e assimétrica positiva, respectivamente.

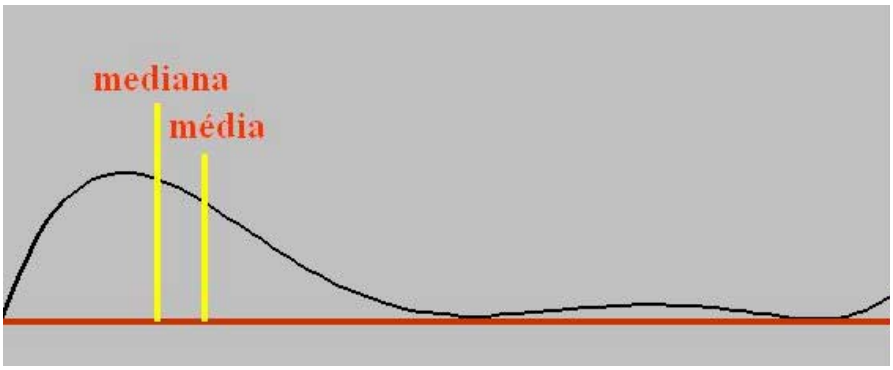


Figura 29– Relação de grandeza entre a média e a mediana numa distribuição de frequências assimétrica negativa.

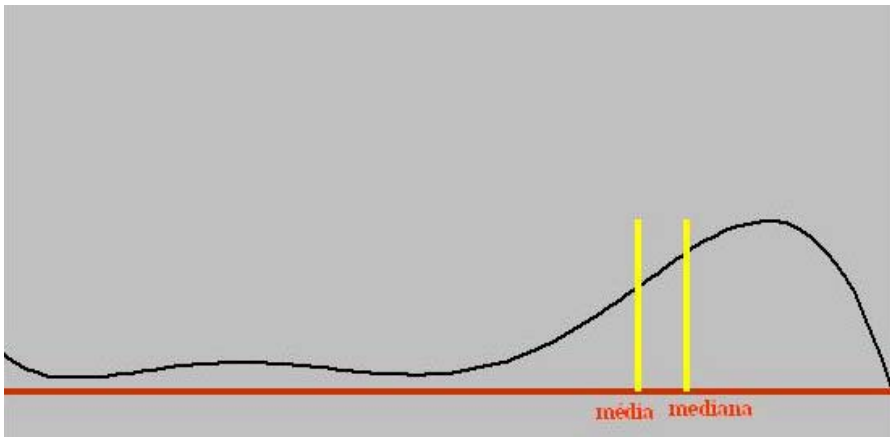


Figura 30– Relação de grandeza entre a média e a mediana numa distribuição de frequências assimétrica positiva.

Assim, uma vez que não conhecemos, à partida, a maior destas medidas de tendência central, definimos um primeiro intervalo de aceitação entre o menor e o maior destes valores (figura 31).

Se a distribuição de frequências fosse uma distribuição Normal, um intervalo natural para aceitação dos tempos de latência seria uma vizinhança da média. Este primeiro intervalo de aceitação (representado na figura 31) irá ocupar as funções da média no nosso algoritmo. É agora necessário definir uma região em torno deste intervalo que seja a região definitiva de aceitação.

A amplitude dos intervalos a adicionar ao intervalo já aceite poderia ser fixa ou variável. Optou-se por uma amplitude variável de modo a que as amplitudes definidas fossem função dos valores apresentados por cada utilizador. Assim, as amplitudes dos intervalos adicionados são, em primeiro lugar, função da média e da mediana dos

valores apresentados pelo utilizador (figura 32) e, portanto, também a amplitude do intervalo final o é. O factores *x* torna o limite inferior do intervalo inicial menor numa

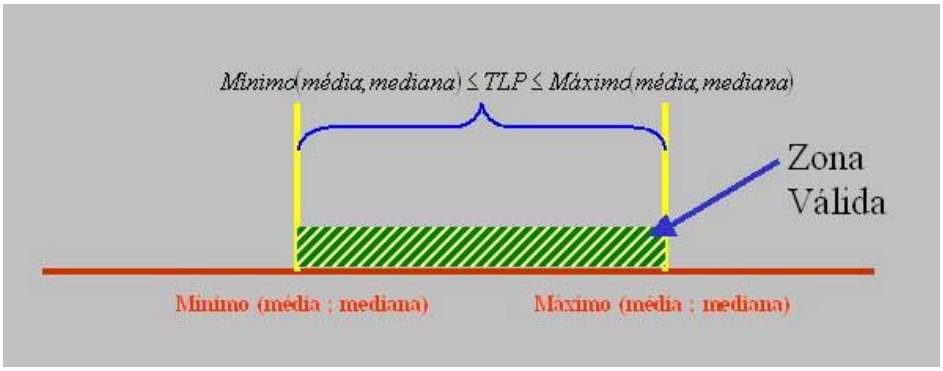


Figura 31 – Um primeiro intervalo de aceitação

determinada percentagem, enquanto que o factor *y* torna o respectivo limite superior maior noutra determinada percentagem.

O desvio padrão é uma medida de dispersão dos dados e, portanto, a amplitude do intervalo final deve ser função deste valor. A expressão $\frac{DesvPadrão}{média}$ transforma o desvio padrão numa fracção da média, representada na forma decimal entre zero e um (já que $\frac{DesvPadrão}{média} \leq 1$ em distribuições próximas da distribuição Normal com todos os valores positivos). Multiplicando esta expressão adicionada de uma unidade pelo extremo superior obtemos o extremo mais uma sua fracção. Multiplicando o extremo inferior por $1 - \frac{DesvPadrão}{média}$ obtemos o extremo menos uma sua fracção. Os testes preliminares mostraram que este intervalo necessita de ser ligeiramente alargado para optimizar os resultados e, portanto, adicionou-se 0,05 (5%) à expressão $\frac{DesvPadrão}{média}$. Este ajuste corresponde a somar 1,05=1+0,05 à expressão no caso do extremo superior e a subtrair a expressão a 0,95=1-(+0,05) no caso do extremo inferior. O valor decidido (5%) resulta da observação experimental, com o objectivo de minimizar o CER, mas pode ser substituído por uma variável, α, que pode ser definida pelo administrador, complicando os níveis de decisão existentes, mas possibilitando mais opções no que respeita ao equilíbrio FRR/FAR. Assim, os pontos de decisão apresentados nesta dissertação referem-se a $\alpha = 0{,}05$. Existem um número infinito de opções

correspondentes a valores de $\alpha = [0; +\infty[$, embora para valores de α muito grandes a permissibilidade seja tal que o sistema deixa de fazer sentido. É necessária mais investigação para estabelecer o intervalo em que α toma valores com sentido, isto é, com vantagem ao nível da dualidade segurança/conforto de utilização.

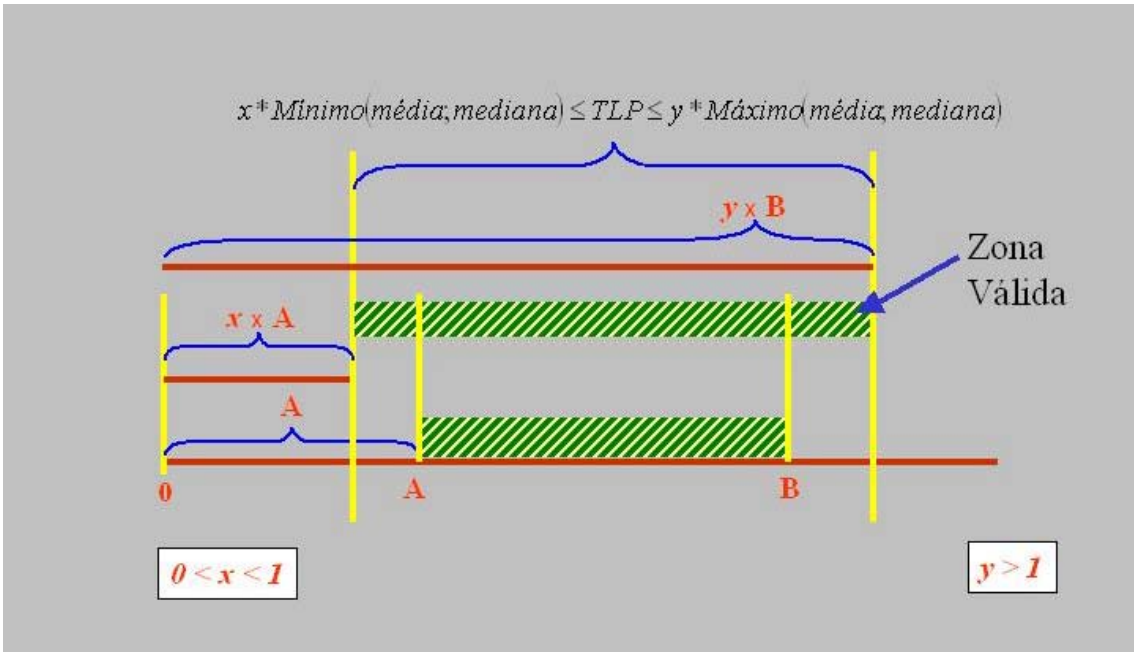


Figura 32 – A amplitude do intervalo como função da média e da mediana dos dados do utilizador.

4.4 – Avaliação do algoritmo

O algoritmo foi desenvolvido partindo dos dados recolhidos de 8 voluntários. Após esta fase, foi necessário testá-lo em maior escala e com pessoas que não estiveram envolvidos no primeiro grupo. Os dados foram recolhidos na Internet (34 utilizadores) e num computador portátil (9 utilizadores), sempre com a mesma sequência de 14 caracteres, com recurso a uma aplicação desenvolvida em Java (figura 33).

Os dados recolhidos no portátil são de pessoas conhecidas do grupo de investigação, 5 homens e 4 mulheres, conhecedores da metodologia utilizada. Através das figuras 34 e 35 pode comparar-se a relação homens/mulheres neste grupo com a distribuição mundial que é, em teoria e a longo prazo, a população alvo desta tecnologia.

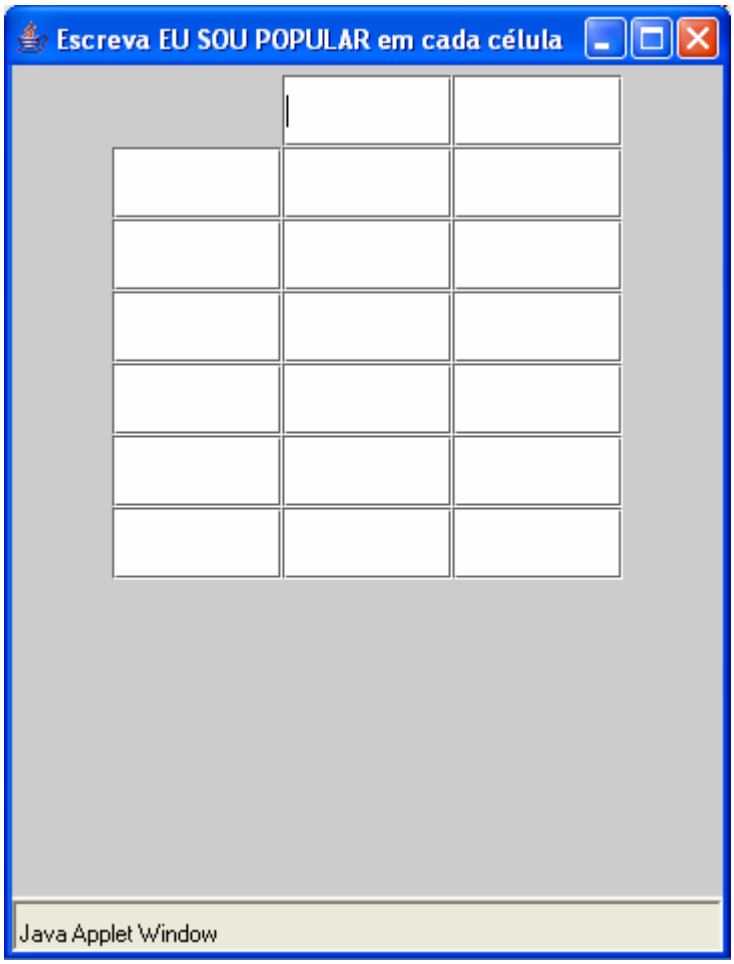


Figura 33 – Janela de introdução da frase passe para captura dos dados biométricos

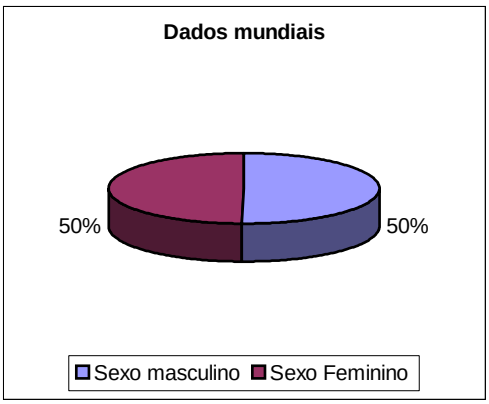


Figura 34 – Distribuição Homens/mulheres na população mundial. Fonte: Nações Unidas.

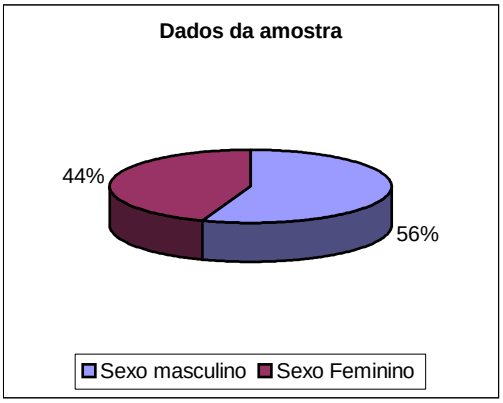


Figura 35 Distribuição Homens/mulheres no grupo de utilizadores conhecidos.

Este grupo representa os utilizadores legítimos que serão atacados por todos os outros voluntários e não estavam autorizados a participar via Internet uma vez que facilmente passariam por eles próprios quando o ataque fosse feito. Não foram colocados quaisquer outros entraves à participação via Internet. Os restantes utilizadores (voluntários anónimos que participaram via Internet) registar-se-iam várias vezes no sistema (estabelecendo uma FRR) e tentariam passar por um dos utilizadores do grupo controlado.

Não foram armazenados quaisquer dados relativos a um utilizador se fossem introduzidos menos do que 16 entradas. Isto permitiu ao sistema utilizar as primeiras 12 entradas para estabelecer um padrão e as restantes para calcular a FRR, considerando-as como tentativas de entrada no sistema. No total, estiveram envolvidos no cálculo da FRR 43 utilizadores, representando 426 tentativas de entrada no sistema.

Cada um dos 9 utilizadores do grupo controlado tinha vários padrões, recolhidos ao longo de 4 semanas, resultantes de diversas entradas no sistema. De cada vez que o utilizador introduzia a sua palavra/frase passe os dados eram armazenados, substituindo os tempos mais antigos. Cada padrão foi atacado por todos os outros utilizadores. Isto resultou em 258 padrões atacados e em 187.545 tentativas ilegítimas de entrada no sistema.

Nas biometrias físicas o utilizador apresenta a mesma característica física em diferentes posições e/ou condições, mas é a mesma característica. Nas biometrias comportamentais o que é apresentado, de cada vez que é tentado o acesso, é um

comportamento diferente, por isso calculou-se a FAR e a FRR considerando o número de tentativas, não de utilizadores.

Para aferir a evolução na precisão obtida com este algoritmo foi calculada a FAR e a FRR do algoritmo de Joyce e Gupta [Joyce, 1990] utilizando a amostragem acima referida e o ponto de decisão por eles publicado – 60% dos tempos de latência têm que cumprir os requisitos para que o utilizador seja aceite. Seguindo idêntico princípio, seleccionou-se o ponto de decisão do algoritmo em avaliação, correspondendo a 60% da máxima classificação possível ou seja, para uma sequência de 14 caracteres $(13\times1.5+1)\times60\% = 20.5\times60\% = 12.3$. Atendendo a que no algoritmo em avaliação o ponto de decisão tem que ser múltiplo de ½, o utilizador será aceite se o seu padrão tiver uma soma A maior do que 12.

Os resultados globais, incluindo a FAR e a FRR obtidos com a utilização do algoritmo em avaliação, bem como os valores equivalentes para um nível de decisão de 60% obtidos a partir do algoritmo de Joyce e Gupta, estão apresentados nas figuras 36, 37 e 38. As figuras 36 e 37 referem-se a valores absolutos (em tentativas) e a figura 38 refere-se a taxas de erro. Na figura 38 estão ainda incluídas as curvas de regressão polinomial da FRR e da FAR para determinar um CER, ao qual corresponde o valor 5,58%.

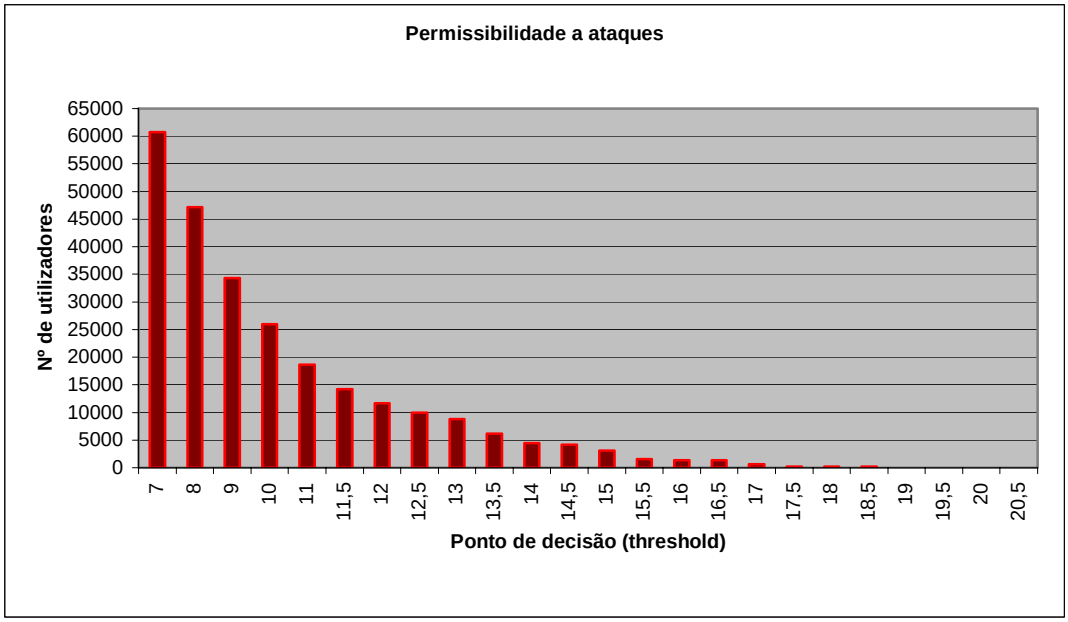


Figura 36– Número de utilizadores ilegítimos com registo bem sucedido, por ponto de decisão.

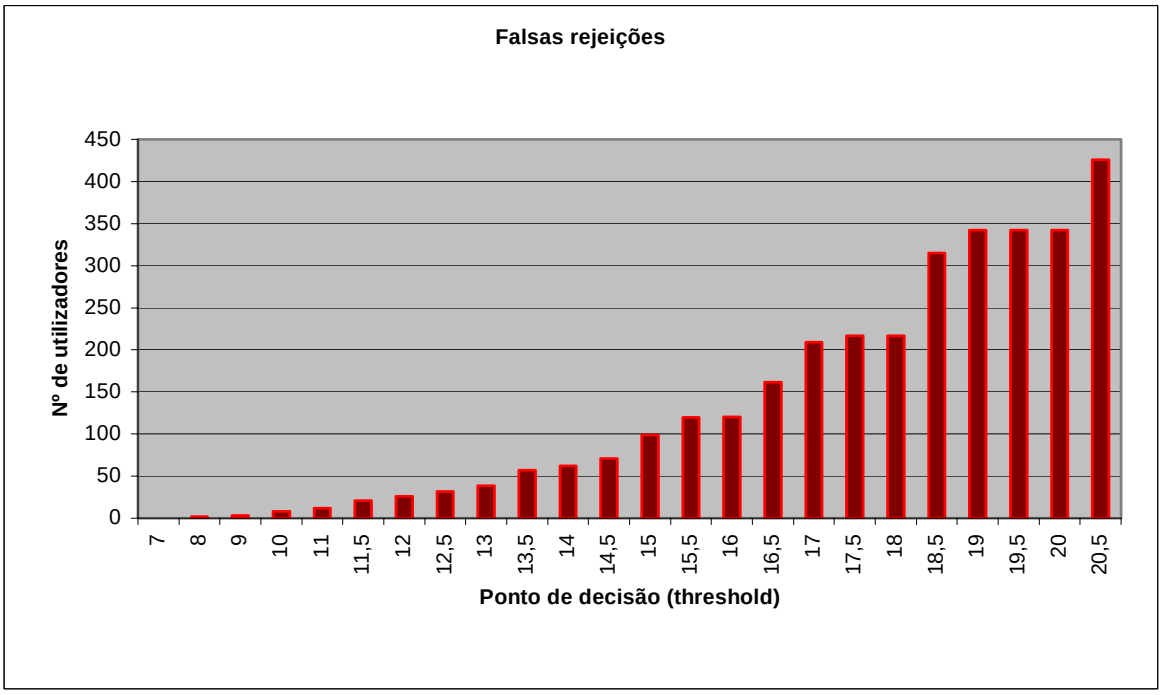


Figura 37 – Número de utilizadores legítimos recusados pelo sistema, por ponto de decisão.

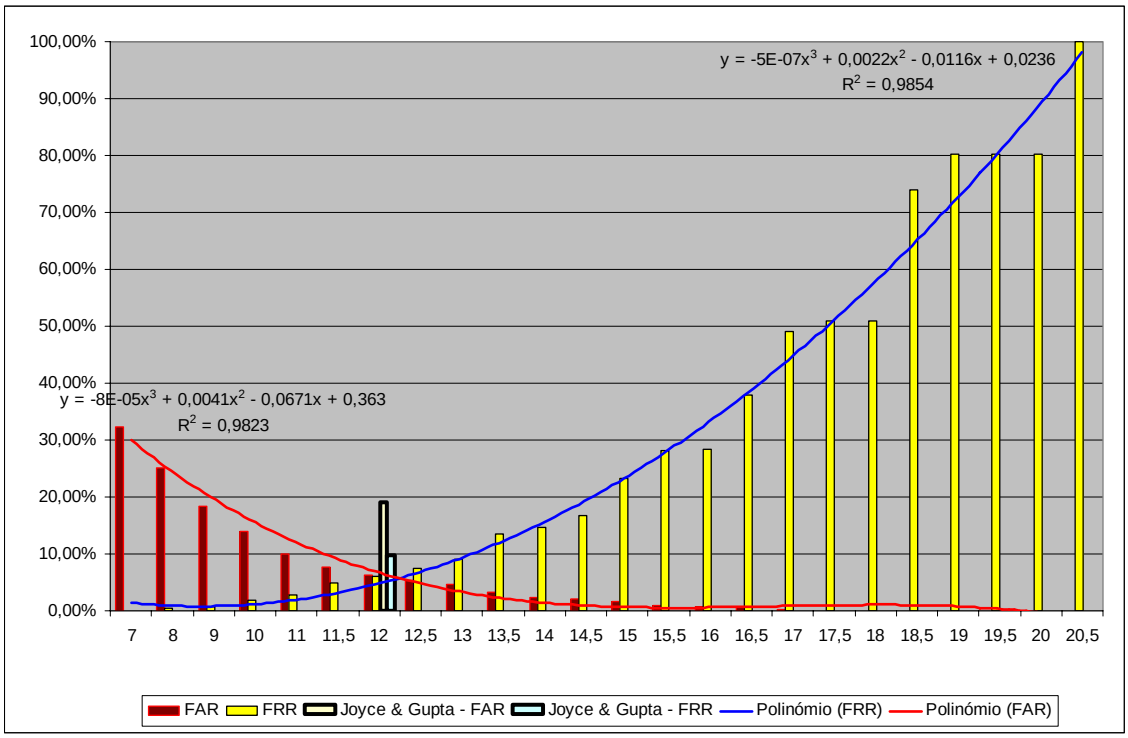


Figura 38 – Resultados globais (%).

Os vários pontos de decisão permitem ao administrador do sistema, ou a um IDS, obter uma FAR próxima de 0%, uma FRR próxima de 0% ou encontrar um equilíbrio, algures no meio, de acordo com as necessidades de segurança do contexto.

Outra possibilidade é estabelecer o ponto de decisão de acordo com a qualidade da palavra/frase passe, isto é, uma sequência maior de caracteres que inclua letras, números e símbolos teria um ponto de decisão mais baixo (mais permissivo) do que uma sequência pequena ou constituída apenas por letras ou apenas por números. Este processo, seria uma adaptação ao facto de uma palavra secreta que se mantenha secreta corresponder a uma diminuição dramática da FAR, como veremos adiante. Assim, é possível diminuir o nível de exigência do algoritmo a utilizadores com sequências complexas, mantendo o nível de segurança pretendido. Este factor pode até servir de estímulo para que os utilizadores abandonem as palavras/frases passe simples, contribuindo a médio prazo para um aumento no nível global de segurança.

Como se pode observar facilmente, o algoritmo aqui proposto evidencia uma melhoria de desempenho importante face aos resultados obtidos com o algoritmo de Joyce e Gupta.

É importante ainda notar que estes valores de FAR e CER são considerados depois de haver uma quebra de segurança, isto é, a palavra passe, que devia ser secreta, passa ser pública. Se a palavra passe, de no máximo *x* caracteres, é considerada secreta, então temos:

.

$$FAR_{palavra\_\text{sec}reta} = \frac{1}{Número\_de\_palavras\_possíveis\_com\_x\_caracteres} \times FAR_{palavra\_pública}$$

Para o caso concreto da palavra testada, 14 caracteres alfabéticos (incluindo o espaço), o número de palavras/frases passe possíveis num alfabeto com 27 símbolos mais o espaço (teclado português convencional) é $28^{14} = 1.82 \times 10^{20}$. Temos então a força do número de possibilidades existente aliada à impossibilidade de fazer um ataque simplesmente por força bruta, uma vez que a componente biométrica o impede.

Os valores apresentados correspondem a valores médios. No entanto, foi possível observar resultados de FRR e de FAR bastante diferentes nos noves utilizadores controlados. Se é certo que nenhum dos resultados pode ser considerado mau, é certo que alguns estarão mais protegidos por este sistema do que outros.

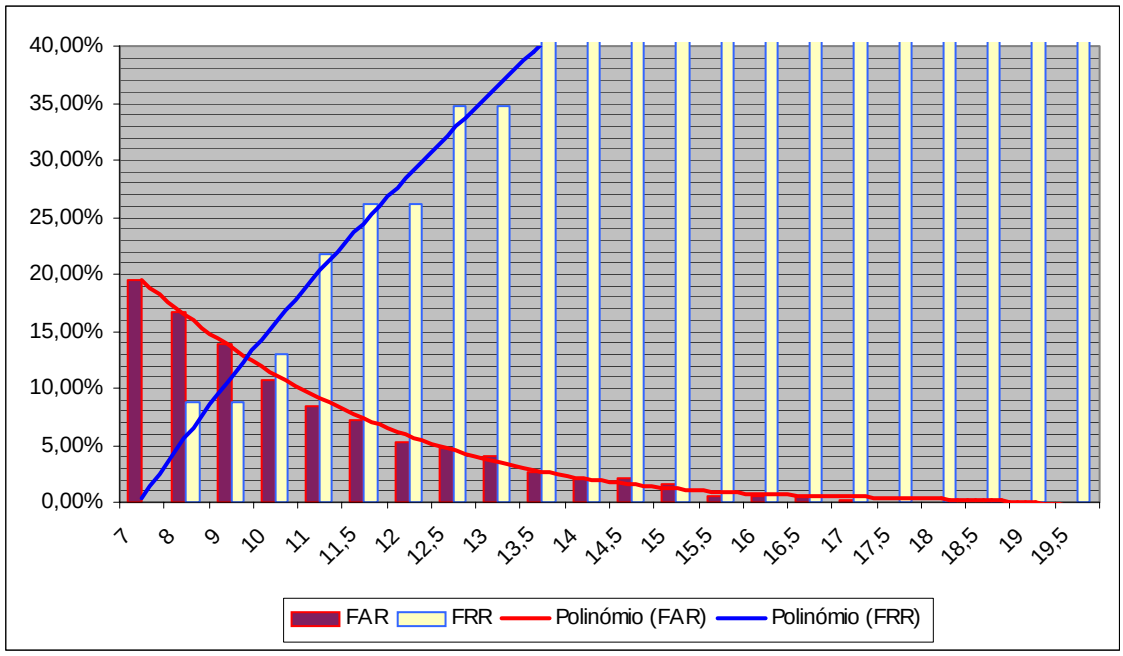


Figura 39 – FAR vs FRR e CER do utilizador com CER mais alto.

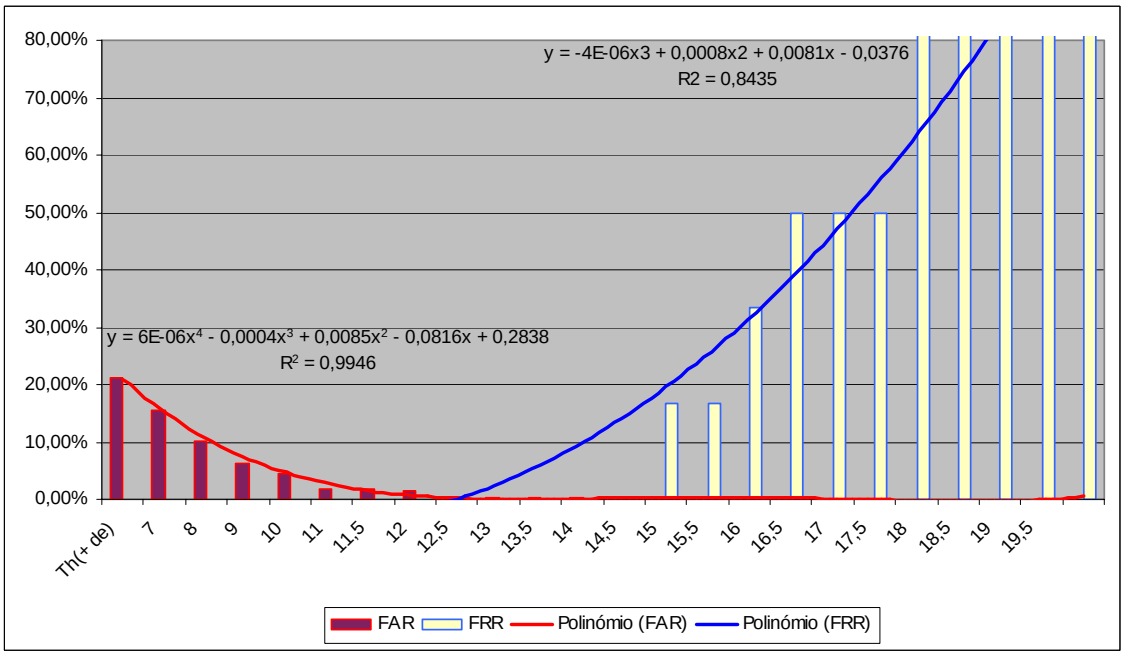


Figura 40 - FAR vs FRR e CER do utilizador com CER mais baixo.

As figuras 39 e 40 apresentam, respectivamente, os resultados do utilizador com CER mais alta e do utilizador com o CER mais baixa. A figura 41 mostra os CER dos 9 utilizadores e, da sua análise, pode-se concluir que os valores apresentados pelos

utilizadores correspondentes aos gráficos das figuras 39 e 40 são excepções à regra e estão muito longe do normal.

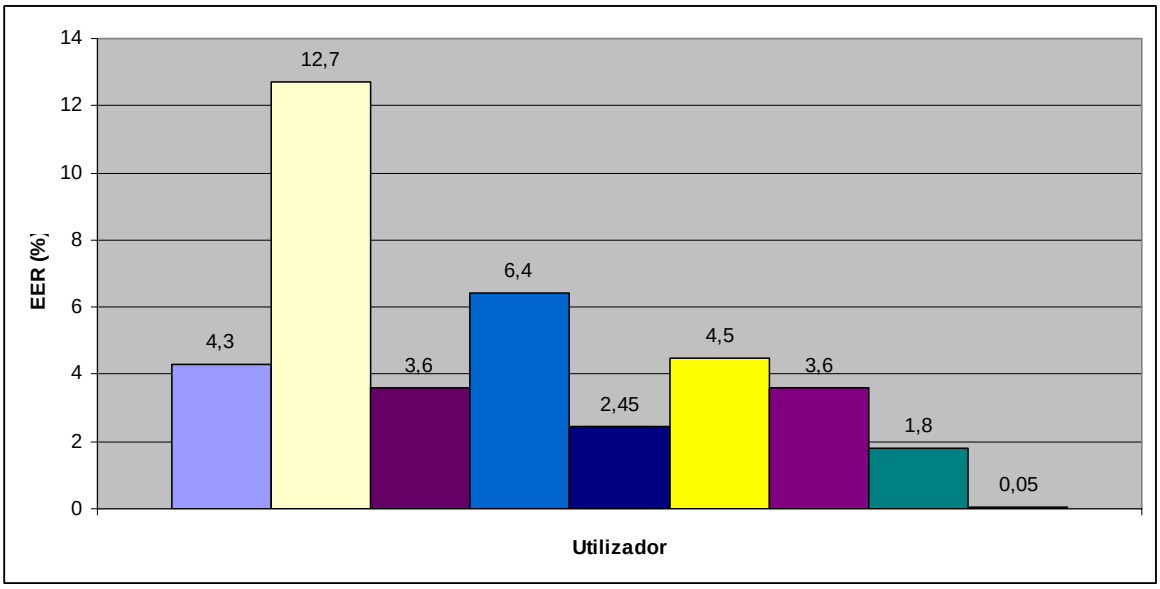


Figura 41 – CER dos utilizadores do grupo controlado.

5 – Trabalho Futuro

De uma forma simples, podemos afirmar que qualquer técnica biométrica assenta na recolha de um conjunto de características próprias de um indivíduo, sendo a autenticação o resultado (positivo ou negativo) da comparação dessas características com um padrão armazenado. Se a recolha deve ser um processo fiável, ao mesmo nível de exigência deve estar a segurança do armazenamento e da operação de comparação. Os SmartCards garantem hoje esse requisito de armazenamento e, com a capacidade de processamento actual, podem ainda efectuar a comparação de padrões, naturalmente com algumas limitações, mas dentro de um ambiente bastante seguro. A tecnologia Java Card, como uma das mais amadurecidas, permite explorar níveis de programação já bastante elevados e garante a portabilidade do processo entre Sistemas Operativos.

Do ponto de vista de sistema existe uma limitação imposta pela separação física entre o subsistema biométrico e o subsistema de suporte ao SmartCard. Esta arquitectura introduz algumas vulnerabilidades na comunicação, que poderão ser ultrapassadas recorrendo a técnicas de encriptação. Contudo, numa evolução previsível, é natural que os leitores biométricos (neste caso o teclado) venham, de uma forma generalizada, a incorporar leitores de SmartCards, conferindo ao conjunto interessantes capacidades de identificação e autenticação. Do ponto de vista da tecnologia é, assim, necessária a criação de uma camada de interface no SmartCard.

Numa política de segurança global, será ainda interessante garantir a integração da função do identificador/autenticador com outros eventuais serviços de segurança existentes num sistema de informação. Esta integração poderá ser feita eficientemente utilizando o protocolo LDAP (Lightweight Directory Access Protocol), bastante utilizado para acesso a repositórios de informação estruturada, como é o caso dos servidores de certificados nas infra-estruturas de chave pública, ou do próprio *registry* que o Windows implementa. Este protocolo, que pode ser seguro quando associado à criptografia, permite interoperabilidade entre sistemas operativos diferentes permitindo, por exemplo, uma autenticação única para Windows e Linux [Swanson et al. 2002]**.** Estes repositórios aparecem então como elo de ligação entre subsistemas independentes de identificação/autenticação e serviços de segurança que exigem essa função.

Quanto ao aperfeiçoamento do algoritmo apresentado, exige um estudo do seu desempenho em palavras/frases passe de dimensões menores, com e sem semântica. O

estudo do desempenho deste algoritmo em *Personal Identification Numbers* (PIN) é fundamental para compreender a sua aplicabilidade em sistemas de segurança largamente disseminados, tais como os sistemas de pagamento electrónico ou as combinações digitais de abertura de cofres. Os resultados obtidos poderão fornecer pistas para o ajuste de condições que optimizem o algoritmo. Uma primeira abordagem deverá substituir os tempos de latência estudados pelos tempos que decorrem entre premir uma tecla e largar a tecla e entre largar a tecla e premir a seguinte. É de esperar que os resultados agora obtidos para uma sequência de 14 caracteres possam, então, ser obtidos para uma sequência de apenas 7 caracteres.

Uma consequência natural dos resultados já obtidos é a implementação do algoritmo. Diversos contextos podem beneficiar desta tecnologia, embora com abordagens diferentes: acesso a conteúdos *Web* protegidos por palavra passe, identificação de utilizadores em postos de trabalho e computadores pessoais (nesta situação o algoritmo tem que ser implementado ao nível do sistema operativo) e até mesmo para reconhecer proprietários legítimos de licenças de software de edição de texto, podendo as licenças pessoais substituir as actuais licenças por posto de trabalho.

Conclusões

Cada uma das técnicas de identificação e autenticação existentes tem, naturalmente, virtudes e inconvenientes, mas atendendo à sua potencial mais valia, as técnicas chamadas biométricas, que procuram utilizar características do indivíduo, têm vindo a evidenciar uma notável evolução. No entanto, a adopção destas tecnologias é travada pela desconfiança dos utilizadores quanto à utilização da sua informação privada e pelo receio de agressões à integridade física, por parte de algumas dessas tecnologias.

O primeiro grupo de limitações é suavizado recorrendo a uma tecnologia de armazenamento e processamento da informação biométrica em *SmartCards*, sob controlo do próprio utilizador. Já o segundo é difícil de garantir nas biometrias físicas devido à natureza intrusiva dos leitores das características biométricas. Os resultados obtidos mostram que é possível utilizar a dinâmica de digitação como uma tecnologia confiável, colaborativa ou furtiva, não intrusiva e de fácil utilização. Além disso, esta biometria pode ser utilizada em pequenos dispositivos, uma vez que o algoritmo apresentado não é exigente do ponto de vista computacional.

No âmbito desta dissertação desenvolveu-se um novo algoritmo de *Keystroke Dynamics*, procurando-se optimizar o modelo estatístico por forma a melhorar a sua precisão e ao mesmo tempo gerar um algoritmo computacionalmente simples que garanta a possibilidade da sua execução nos limitados recursos de um *SmartCard*. Uma vez descoberta a palavra/frase passe, o algoritmo apresentado tem um CER de 5,58%. Mas isto significa que, sem um esforço extra, é 94,42% mais difícil alguém fazer-se passar por um utilizador legítimo. Se a palavra/frase passe é mantida secreta, o CER tem um valor competitivo, quando comparado com as tecnologias biométricas mais precisas.

O comportamento humano muda com o tempo. Mas este algoritmo usa um padrão armazenado de forma dinâmica, isto é, que evolui com as mudanças do utilizador, uma vez que só os últimos 12 registos bem sucedidos são armazenados.

Os resultados obtidos utilizando o algoritmo de Joyce e Gupta com os dados recolhidos mostram que existe, de facto, uma evolução. Ainda assim, são necessários mais estudos para avaliar o desempenho do algoritmo quando utilizado com sequências de caracteres mais pequenas ou com códigos PIN. De notar, contudo, que a ausência de

benchmarks para este efeito limita a generalização das conclusões. No entanto, pela experiência obtida e os casos avaliados, é firme a convicção do autor de que o algoritmo proposto é de enorme utilidade para a implementação de medidas preventivas, relativas à identificação e autenticação, nas políticas de segurança para os Sistemas de Informação.

Índice de figuras

Índice de tabelas

Bibliografia

Berta, I. Z. e Mann, Z. A.: *SmartCards – Present and Future*, Híradástechnika, Journal on C5, 2000.

Biometric Watch$_{TM}$: *Colombian Bank Implements Biometric ATM's*, Vol. 2, n. º 10, 2004

Chen, Z., *Java Card Technology for SmartCards*, Addison Wesley, U.S.A., 2000.

CNPD – Comissão Nacional de Protecção de Dados: *Princípios sobre a utilização de dados biométricos no âmbito do controlo de acessos e de assiduidade*, www.cnpd.pt (Novembro, 2004), 2004.

Davies, S.: *Touching Big Brother – How biometric technology will fuse flesh and machine*, Information Technology & People, Vol 7, No. 4, 1994.

Epaynews: *payment news and resource center*, www.epaynews.com/poll/index.html, (Dezembro, 2004), 2004.

IBG, International Biometric Group: *The Biometric Industry: One Year After 9/11*, http://www.biometricgroup.com/reports/public/reports/9-11.html, (Novembro, 2004), 2003.

ISO – International Organization for Standardization, *ISO/IEC 7816*, http://www.iso.org/iso/en/CatalogueListPage.CatalogueList?COMMID=64&scopelist= (Outubro, 2004)

Jain, A., Hong, L e Pankanti, S.: *Biometric Identification*, Comunications of the ACM, Vol. 43, No. 2, 2000.

Joyce, R e Gupta, G.: *Identity authorization based on keystroke latencies*, Communications of the ACM, 33(2): 168-176, 1990.

Kumar, A., Wong, D. C. M., Shen, H. C. e Jain, A. K.: *Personal Verification using Palmprint and Hand Geometry Biometric*, Proc. of 4th Int'l Conf. on Audio- and Video-Based Biometric Person Authentication (AVBPA), Guildford, UK, 2003.

Lei 67/98 de 26 de Outubro, *Lei da Protecção de Dados Pessoais*, disponível em www.cnpd.pt/Leis/lei_6798.htm (Novembro, 2004), 1998

Liu, S. e Silverman, M.: *A Practical Guide to Biometric Security Technology*, IEEE Computer Society, www.computer.org/itpro/homepage/Jan_Feb/security3.htm (Dezembro de 2002), 2001.

Luis-García, R., Alberola-López, C., Aghzout, O. e Ruiz-Alzola, J.: *Biometric Identification systems*, Signal Processing, Elsevier North-Holland, Inc, Amsterdam, The Netherlands, volume 83, 2539-2557, 2003.

Magalhães, P. S. e Santos, H. D.: *Biometria e Autenticação*, Actas da 4ª Conferência da Associação Portuguesa de Sistemas de Informação". Porto. Portugal. 15-17/10/2003 (edição em CD-ROM: ISBN 97 2-9354-42-1)

Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L. e Jain, A. K.: *FVC2002: Second Fingerprint Verification Competition*, Proceedings of the International Conference on Pattern Recognition – ICPR2002, 2002

Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L. e Jain, A. K.: *FVC2000: Fingerprint Verification Competition*, relatório técnico, http://bias.csr.unibo.it/fvc2000 (Junho 2004), 2001

Maio, D., Maltoni, D., Cappelli, R., Wayman, J.L. e Jain, A. K.: *FVC2004: Third Fingerprint Verification Competition*, Proceedings of the International Conference on Biometric Authentication – ICBA, Hong Kong, 2004

Maltoni, D.Maio, D., Jain, A. K. e Prabhakar, S.: *Handbook of fingerprint recognition*, Springer, New York, 2003.

Markowitz, J.: *Voice Biometrics*, Comunications of the ACM, Vol. 43, No. 9, 2000.

Monrose, F e Rubin, A. D.: *Authentication via Keystroke Dynamics*, Proceedings of the Fourth ACM Conference on Computer and Communication Security, Zurich, Switzerland, Abril, 1997.

Monrose, F e Rubin, A. D.: *Keystroke Dynamics as a Biometric for Authentication*, Future Generation Computing Systems (FGCS) Journal: Security on the Web. Março de 2000.

Monrose, F, Reiter, M. K. e Wetzel, S.: *Password Hardening based on Keystroke Dynamics*, International Journal of Information Security, 2001

Ord, T e Furnell, S. M.: *User authentication for keypad-based devices using keystroke analysis*, Proceedings of the Second International Network Conference – INC 2000, Plymouth, UK, 2000

Peacock, A., Ke, X. and Wilkerson, M: *Typing Patterns: A Key to User Identification*, IEEE Security and Privacy, September/October 2004.

Phillips, P. J., Grother, P., Micheals, R. J., Blackburn, D. M., Tabassi, M. e Bone, M.: *Face Recognition Vendor Test 2002: Evaluation Report*, www.frvt.org, (Abril 2003), 2003

Privacy International, Statewatch e European Digital Rigts: *An Open Letter to the ICAO A second report on 'Towards an International Infrastructure for Surveillance of Movement'*, 30 de Março, 2004

Privacy International, Statewatch e European Digital Rigts: *An Open Letter to the European Parliament on Biometric Registration of All EU Citizens and Residents*, 30 de Novembro, 2004.

Putte, T. e Keuning, J.: *Biometrical fingerprint recognition: don´t get your fingers burned*, Proceedings of IFIP TC8/WG8.8 Forth Working Conference on SmartCard Research and Advanced Applications, Kluwer Academic Publishers (2000), 289-303.

Poh, N. e Korczak, J.: *Hybrid Biometric Person Authentication Using Face and Voice Features*, Proceedings of the Third International Conference, Audio- and Video--based Biometric Person Authentication AVBPA 2001, Halmstad, Sweden, 2001, 348-353.

Ross, A.: *A Prototype Hand Geometry-based Verification System*, M. S. Project Report, http://biometrics.cse.msu.edu/RossHand_MS99.pdf (Setembro, 2004), 1999

Swanson, C. E Lung, M.: *OpenLDAP everywhere*, Linux Journal, Vol. 2002, No. 104, 2002.

Thian, N.: *Biometric Authentication System*, Tese de mestrado, USM, Penang, Malásia, http://hydria.u-strasbg.fr/~norman/BAS/publications.htm (Fevereiro 2003), 2001.

U. S. Department of Homeland Security: *Machine-Readable Passport Requirement*, Press Release, USA, 22 de Outubro, 2004.

U. S. Department of State: *Extension of Requirement for Biometric Passport Issuance by Visa Waiver Program Countries*, Press Statement 2004/886, Washington, DC, 10 de Agosto, 2004.

VISA Europe: *VISA Europe Region*, comunicado à imprensa, acedido em http://www.visaeurope.com/pressandmedia/visaeuropeanunionregion.html (Setembro, 2004), 2004.

Wang, Y., Tan, T. e Jain, A. K., *Combining Face and Iris Biometrics for Identity Verification*, Proc. Of 4th Int'l Conf. On Audio- and Video-Based Biometric Person Authentication (AVBPA), Guildford, UK, 2003.